UNIVERSITY OF MINNESOTA RECEIVED NOV 9 1942 LIBRARY

The Journal of Experimental Education

A periodical report of scientific investigations relating to child development, curriculum, learning, teaching, supervision, measurements, statistics, and experimental techniques.

Volume XI | SEPTEMBER, 1942 | Number 1

LEARNING, TEACHING, SUPERVISION

CONTENTS

PAGE

The Effect of Diagraming on Certain Skills in English Composition: *James Reese Stewart* ---- 1

An Evaluation of an Individual Method and a Group Method of Teaching College Freshmen the Mechanics of English Composition: *Mark Karp* ---- 9

The Nature and Causes of Regressive Movements in Reading: *Evalyn Bayle* ---- 16

An Experimental Study of the Relationship of Peripheral Perception to Factors in Reading: *Cyrus W. La Grone, Jr.* ---- 37

Efficiency in Learning: An Experimental Study of Learning by Memorization, by a Verbal Rule, and by Non-Verbal Clues to Understanding: *J. C. Hewson* 50

Factors Associated with the Achievement of High School Pupils of Superior Intelligence: *John W. Musselman* ---- 53

An Experimental Study of Two Methods of Teaching Chemistry in Senior High School: *William E. Ward* ---- 69

The Reliability and Permanence of Vocational Interests of Adolescents: *Katherine van Fridagh Taylor* ---- 81

The Relationship Between Performances on Examinations and Emotional Responses: *Wm. H. Waite* ---- 88

$5.00 A YEAR | PUBLISHED QUARTERLY | $1.50 A COPY

Edited and Published by A. S. Barr, Professor of Education, University of Wisconsin, Madison, Wisconsin.

Entered as second-class matter October 17, 1938 at the post office at Madison, Wisconsin, under the act of March 3, 1879.

EDITORIAL BOARD

A. S. Barr, Chairman, Professor of Education, University of Wisconsin, Madison, Wis.

Carter V. Good, Professor of Education, University of Cincinnati, Cincinnati, Ohio. Editorially responsible for materials on learning, teaching and supervision, published each September.

J. Wayne Wrightstone, Assistant Director, Bureau of Reference, Research and Statistics, Board of Education of the City of New York, 110 Livingston Street, Brooklyn, New York. Editorially responsible for materials on curriculum construction, published each June.

Edward E. Cureton, Senior Educational Statistician, U. S. Office of Education, Washington, D. C. Editorially responsible for materials on measurements, statistics, and methods of experimental research, published each March.

Arthur T. Jersild, Professor of Education, Teachers College, Columbia University, New York City. Editorially responsible for materials on child welfare, guidance, and development, published each December.

CONTRIBUTING EDITORS

Gilbert L. Betts, Supervisor of Graduate Research in Education, Colorado State College, Fort Collins, Colorado.

William A. Brownell, Professor of Educational Psychology, Duke University, Durham, North Carolina.

Leo J. Brueckner, Professor of Education, University of Minnesota, Minneapolis, Minnesota.

Oscar K. Buros, Associate Professor of Education, Rutgers University, New Brunswick, New Jersey.

Otis W. Caldwell, General Secretary. The American Association for the Advancement of Science, Boyce Thompson Institute for Plant Research, Yonkers, New York.

Leslie L. Chisholm, Associate Professor of Education, State College of Washington, Pullman, Washington.

Herbert S. Conrad, Associate Professor of Psychology, College of Agriculture, and Research Associate, Institute of Child Welfare, University of California, Berkeley, California.

Stephen M. Corey, Professor of Educational Psychology, University of Chicago, Chicago, Illinois.

Robert A. Davis, Professor of Education, Director of Bureau of Educational Research, University of Colorado, Boulder, Colorado.

Harl R. Douglass, Director of College of Education, University of Colorado, Boulder, Colorado.

Jack W. Dunlap, Associate Professor of Educational Psychology, University of Rochester, Rochester, New York.

Harold A. Edgerton, Assistant Professor of Psychology, Ohio State University, Columbus, Ohio.

Alvin C. Eurich, Professor of Education, Stanford University, Stanford University, California.

John C. Flanagan, Assistant Chief, Research Section, Medical Division, Office, Chief of the Air Corps, War Department, Washington, D. C.

Kai Jensen, Associate Professor of Education, University of Wisconsin, Madison, Wisconsin.

Harold E. Jones, Professor of Psychology, Director of Research, Institute of Child Welfare, University of California, Berkeley, California.

Noel Keys, Professor of Education, University of California, Berkeley, California.

Edward A. Lincoln, Consulting Psychologist, Halifax, Massachusetts.

T. E. Newland, Pennsylvania State Department of Education, Chief of Special Education, Harrisburg, Pennsylvania.

C. W. Odell, Associate Professor of Education, University of Illinois, Urbana, Illinois.

Willard C. Olson, Professor of Education, Director of Research in Child Development. University of Michigan, Ann Arbor, Michigan.

W. E. Peik, Dean and Professor of Education, University of Minnesota, Minneapolis, Minnesota.

S. L. Pressey, Professor of Educational Psychology, Ohio State University, Columbus, Ohio.

Clarence E. Ragsdale, Associate Professor of Education, University of Wisconsin, Madison, Wisconsin.

H. H. Remmers, Director, Division of Educational Reference, Professor of Education and Psychology, Purdue University, Lafayette, Indiana.

Henry D. Rinsland, Professor of Education and Director of Educational Research, The University of Oklahoma, Norman, Oklahoma.

Robert T. Rock, Jr., Professor of Psychology, Head of Dept. of Psychology, Graduate School, Fordham University, New York City.

G. M. Ruch, Chief of Research and Statistical Service, U. S. Office of Education, Washington, D. C.

P. J. Rulon, Assistant Professor of Education, Graduate School of Education, Harvard University, Cambridge, Massachusetts.

Douglas E. Scates, Associate Professor of Education, Duke University, Durham, North Carolina.

David Segel, Educational Consultant, Specialist in Tests and Measurements, Federal Security Agency, U. S. Office of Education, Washington, D. C.

Paul W. Terry, Professor of Educational Psychology, University of Alabama, University, Alabama.

Helen Thompson, Clinic of Child Development, Research Associate, Yale University, New Haven, Connecticut.

Robert L. Thorndike, Associate Professor of Education, Teachers College, Columbia University, New York City.

Herbert A. Toops, Professor of Psychology, Ohio State University, Columbus, Ohio.

T. L. Torgerson, Professor of Education, University of Wisconsin, Madison, Wisconsin.

Helen M. Walker, Professor of Education, Teachers College, Columbia University, New York City.

Beth L. Wellman, Professor of Psychology, Child Welfare Research Station, State University of Iowa, Iowa City, Iowa.

Guy M. Wilson, Professor of Education, Boston University, Boston, Massachusetts.

Paul A. Witty, Professor of Education, Director of Psycho-Educational Clinic, School of Education, Northwestern University, Evanston, Illinois.

Ernest R. Wood, Professor of Education, New York University, New York City.

DEMOCRAT PRINTING COMPANY
MADISON, WISCONSIN

Journal of Experimental Education

Volume XI | September, 1942 | Number 1

THE EFFECT OF DIAGRAMING ON CERTAIN SKILLS IN ENGLISH COMPOSITION[1]

James Reese Stewart
University of Cincinnati

The time is rapidly passing when instructional materials and teaching methods may be based on tradition and personal opinion alone, with no basis of scientific evidence as to their worth. Teachers and supervisors are expected to know considerable about not only what is to be taught in a given subject and at what grade levels, but also the most effective methods to be used in teaching the subject.

It is believed that teachers in the field of English have subjected their teaching, the methods, and the materials to more minute and severe criticism than have teachers of other subjects in the schools (4:v). According to Fries (5:vii), "the National Council of Teachers of English has come more and more to realize that in order to avoid the futile and even harmful teaching practices which have resulted from ignorance, satisfactory programs for teaching the English language in the schools must rest upon essentially sound views of language and a knowledge of the facts." This is perhaps merely another way of expressing the conviction that research and the evidence produced by it are essential to the improvement of instruction in the English program.

By 1941, almost every phase of the language-composition area had been subjected to analytical study. Like other subjects in the curriculum, the various phases of English began to be subjected to careful investigation shortly after 1900, and controversy on many issues has continued through the past forty years.

The Problem

Background of the Problem.—Diagraming of sentences was widely used as a method of teaching in the language-composition area of English from the time of early records in educational methods well into the 1920's. Then, for no reason based on research, it was discarded almost entirely from the English textbooks for use in elementary and high schools. Within the past five or six years simple diagraming has returned for no more reason apparently than it disappeared. Probably because it was traditional, however, some schools used the method straight through these periods of its decline and rise in popularity.

Educators need only to refer to their own memories for evidence that diagraming was commonly used as a method of teaching language composition as late as 1925. However, for objective proof that diagraming was in the past widely used, methods books and textbooks in the field may be consulted. The following statement is representative of what is found: "It was the minute and mechanical parsing of words and the elaborate diagraming and analyzing of sentences that gave grammar a bad name (in the past)" (4:81).

For those who demand evidence that the practice of diagraming sentences is now regaining popularity in the newer textbooks, methods books in the field may be consulted as a first source. The authors quoted above (4:81–83, 242) found it significant to observe that new textbooks in composition and grammar in the three or four years prior to 1939 were frankly returning to the use of simple diagrams to aid pupils in understanding sentence structure. These authors themselves feel that there is no educational malpractice in making a diagram of a sentence, and they go so far as to point out specialized uses of diagraming.

Ward (25:365), in a discussion of sentence diagraming, very definitely advocated the practice. He stated that for advanced classes, if the teacher is adept, a fully analyt-

[1] Doctor's dissertation completed in August, 1941, under the direction of Dr. H. A. Greene at The State University of Iowa.

ical method of diagraming may be successful; but for the ordinary high-school class such a method is likely to defeat its own purpose. Ward felt that only the simplest exhibit of the chief elements of a sentence is advisable. He stated, however, that the pupil who is unable to express in some graphic way the structure of sentences does not understand sentences, and pupils who are trained to chart the structure learn much.

Although Parker (21:147) is very much opposed to teaching parts of speech and like concepts, he sets up a system which he advocates for analyzing sentence patterns. The system as described resembles a written analysis, which is usually considered in the same light as diagraming; but at the same time this system is so complicated that it appears to rival the old-fashioned, elaborate diagraming.

Some other authors do not state specifically that a written or graphic analysis is their intended proposal; such is implied, however, from their discussions (6:156; 23:641).

As a second source of evidence of the return of diagraming, recent English textbooks and workbooks in the Educational Library at the State University of Iowa and the collection of such books owned by a member of the English staff at the University were examined. It was found that the trend is definitely toward inclusion of sentence diagraming to a greater or lesser degree.

A third source of evidence of the present use of diagraming as a teaching method in English composition was found in the replies to an inquiry sent in January, 1941, to superintendents of schools in seven mid-western states. Of the forty-two answers received which indicated the amount of diagraming taught in their respective schools in the past five years, only fourteen signified that no diagraming had been taught; twenty-four indicated that a little diagraming had been taught, and four that much diagraming had been taught during the previous five-year period.

Previous Investigations.—In view of the general belief by educators that the way to evaluate methods and materials of instruction is to apply scientific methods, and, at the same time, the fact that diagraming of sentences in English teaching has declined and recently risen again in popularity, it seems logical to assume that somewhere in the published records, or in the available unpublished records, are many studies which have attempted to evaluate this method of teaching. In the search for objective evidence relative to the value of diagraming sentences as a teaching method, all compilations of investigations done in the field of language composition from 1900 to 1941 known to the writer were consulted (3,7,8,9,11,13,14,15,16, 17,18,19,20,22). In none of these references was mention made of a study directed toward any aspect of the use of diagraming of sentences as a teaching method.

Only two studies were found that attacked any phase of diagraming specifically. These two were unpublished masters' theses at the University of Iowa. A study (1) of the effects of sentence diagraming on English usage and reading comprehension was completed in 1940; the second, a study of the effects of sentence diagraming on English correctness and reading comprehension (2), was being carried on at the same time as the present investigation.

It appears that, although there is a definite trend toward a return to a more formal type of grammar, of which the diagraming of sentences is an important phase, there is little evidence based on significant experimental research to evaluate its worth objectively and experimentally. It is the purpose of this study to make such evidence available.

Statement of the Problem.—The purpose of this investigation was to study the effect of sentence diagraming on certain English skills. In this study, answers to the following specific questions were sought:

1. Does more learning of usage, capitalization skills, and punctuation skills take place as a corollary of an instructional program emphasizing the diagraming of sentences than takes place as a consequence of a program stressing composition?

2. Is diagraming as a method of teaching grammar information more effective than a method emphasizing composition?

3. Does diagraming as a method of teaching sentence structure enable pupils to write more mature, varied, and pleasing sentences than a method emphasizing composition?

Procedure

The effect of sentence diagraming on capitalization, punctuation, usage and sentence structure skills, and grammar information

was attacked by comparing the results of two parallel and systematic eight-week instructional programs using approximately 1000 pupils in two ninth-grade English classes from twenty Iowa, Illinois, and Minnesota schools.[2]

Before an attempt was made to secure these classes for use in the experiment, it was decided to establish three criteria for the acceptance of the pairs of classes to be included in the study: (1) In order to control as nearly as possible the teacher variable, the two classes in each school were to be taught by the same teacher. (2) To secure classes approximately equal in ability, only those schools that made no use of ability grouping in their English classes were to be accepted. (3) Only those schools that had taught little or no diagraming for the past several years were to be enrolled.

Realizing that these qualifications would eliminate many schools, the investigator mailed letters to one hundred school superintendents in seven mid-western states outlining the proposed program and inquiring as to their willingness to participate in it. From the reply cards received signifying willingness to participate in the study, twenty that seemed to meet best the qualifications previously set forth were selected in which to carry out the experiment. By the use of a table of random numbers, one of the two ninth-grade English classes in each participating school was designated as a "control group" and the other as an "experimental group." The program to be used in one set of classes emphasized composition exercises, while the other stressed diagraming exercises.

Two sets of instructional materials were constructed for use in this problem. For the experimental or diagraming classes, lessons making up a forty-eight-page booklet were reproduced in only slightly modified form from *Adventures Wise and Otherwise*, by Walker, Bartels, and Marye (24).[3] This ninth-grade text in workbook form is the first of a series of four "habits and skills" books intended for use throughout four years of high school and is one of the many that are returning frankly to the use of diagraming as a teaching method. Sixty-one pages from the workbook and seven mastery tests that accompany it were reproduced either wholly or in part. With the exception of a few introductory lessons for new units and some mastery tests, each of the forty lessons reproduced for the experimental groups used diagrams as the exclusive method of teaching the grammatical concepts presented. Drill on seventeen such concepts was carried out in the series of lessons. These concepts were: simple subject, compound subject, predicate verb, compound verb, adjective, adjective phrase, adjective clause, adverb, adverbial phrase, adverbial clause, predicate adjective, predicate nominative, direct object, indirect object, simple sentence, compound sentence, and complex sentence. Two punctuation situations were so intimately connected with certain of these concepts that it was thought impossible to separate the two. Therefore, punctuation of transposed clauses and punctuation of compound sentences were included among the concepts presented.

The instructional program for the experimental group was made up of thirty-three lessons that provided practice in use of the nineteen concepts named above; the remaining seven lessons were review lessons over the concepts. In order that all groups of pupils might complete the forty lessons in the specified time, and at the same time some groups not exhaust the materials before the close of the experimental period, five groups of additional exercises were presented at the end of the lesson series. The latter were so designed as to be usable for supplementary materials in cases where the lesson intended for a given school period was not long enough to occupy the allotted time.

For the control or composition groups of ninth-grade pupils, a fifty-two-page booklet of forty lessons was constructed by the investigator to parallel exactly the nineteen concepts studied by the experimental classes. Whereas, in the lessons for the experimental group, diagraming of sentences was emphasized as the teaching method, in the control group the teaching methods stressed were the writing of original sentences and the rewriting of poor sentences. This book was so constructed that *no diagraming* would be done by the pupils in the control classes. The first thirty-three lessons for this group also provided practice in use of the nineteen con-

[2] Alexis and Waverly, Illinois; Glencoe and Tracy, Minnesota; Boone, Britt, Chariton, Decorah, Harlan, Mason City, Monticello, Oskaloosa, Shenandoah, Sibley, Storm Lake, Tama, Villisca, Vinton, Wapello, and Washington, Iowa.

[3] The writer is indebted to Harcourt, Brace and Company for making the study possible by generously permitting the reproduction of these materials.

cepts. These were followed by seven review lessons. In this set of materials also the last five groups were constructed to be used as supplementary exercises throughout the instructional period, if needed.

The instructional materials and initial tests, along with instructions for the administration of each, were mailed to all participating schools during the last week in February and the first week in March, 1941. These materials included for each pupil in the participating classes: an instructional booklet, a *1940 Iowa Every-Pupil Test in English Correctness* (the usage, capitalization, and punctuation sections only to be used), a grammar information test, and a new-type sentence structure test designed by the writer to measure a pupil's ability *to construct* good sentences rather than his ability *to recognize* good sentences.

Since diagraming is concerned with the structure of the individual sentence, it was apparent from first consideration of this problem that a device which would measure the ability of a pupil to construct sentences was necessary as a satisfactory evaluating instrument. None of the existing measuring devices seemed adequate for use in this situation, since all attempted only to measure the pupil's ability to recognize good sentences and made no attempt to measure his ability to construct them. Therefore, an instrument that would evaluate the quality of sentences as written by the pupil was developed. In order to insure each pupil's encountering the same situations, ideas were supplied in a series of short simple statements and the pupils asked to organize them into the most interesting and effective sentences possible. Operations necessary to combine these ideas into effective sentences constituted items in the test. Scoring was made objective through the cooperation of twelve competent English teachers. The reliability and indications of the validity of this test were secured. Correlations between both teachers' judgments and performance, as measured by the *Willing Composition Scale* and scores on the sentence structure test, indicated that the test possessed a relatively high degree of validity (10). At the same time, the within-groups correlation technique yielded a reliability coefficient of .78 for the sentence structure test.

Special directions that accompanied the instructional booklets and tests asked the teachers to administer the tests as soon as possible after they were received and to enter upon the instructional program immediately thereafter. The teachers were directed to continue the instructional program with both groups for forty class periods, using one period for each lesson outlined in the instructional booklets. The teachers were asked to give each group under their instruction the same enthusiastic treatment and the same opportunity and amount of time to learn the materials presented, and to make the lessons for each group as effective as possible by discussing the pupils' exercises with them each day as extensively as time would permit. In order to eliminate as completely as possible all extraneous factors that might affect the results of the experiment, the full class period each day was to be given over to this work regardless of the length of the period.

At the close of the instructional period the final tests included the *1939 Iowa Every-Pupil Test in English Correctness* (usage, capitalization, and punctuation sections only), the same grammar information test used in the initial testing, a second form of the sentence structure test, and a sentence diagraming test.

The initial and final tests were returned to the writer and were scored under his supervision. When the final tests were returned, the teachers in all twenty schools were asked to report the names of pupils absent for more than ten class periods during the course of the experiment. The scores of these pupils on all tests were omitted from the data of the experiment. In addition, due to the nature of the statistical procedure involved, a pupil's score in a given area could not be used unless he had taken both the initial and final tests.

Since the sentence diagraming test was not to be used in the analysis of the data, but was intended solely as a check on whether the instructional program had been properly carried out, these papers were not scored. Each group of the papers for all twenty schools, however, was examined carefully for information that might be revealed with regard to the instructional program.

After all tests were scored, the data of the experiment were analyzed by the methods of analysis of covariance which, by the use of

initial measures, secures the same increase in precision as does the matching or equating of groups. A description of these techniques as they are applied in this problem is given by E. F. Lindquist (12:180–207). His example is, with one modification, identical to the procedure followed in the present analysis. Since intact classes with no reorganization were enrolled in this study, there was a variation in the size of class within schools. An adjusted number equal to the average, rounded to the nearest even number, of the two classes was computed for each school. This adjusted class number was then used to compute an adjusted total score, the product of the adjusted class number and the class mean, for each class. These adjusted values were substituted for the actual numbers and totals in the subsequent analysis.

Results

The initial and final test scores on each of the five measures (usage, capitalization, punctuation, grammar information, and sentence structure) were carefully and critically analyzed.

Each of the following five tables deals with one of the five aspects of the problem. Each shows the analysis of covariance for the initial and criterion class mean scores in a specific area tested. In order to facilitate interpretation, the general means for the criterion tests, adjusted to allow for chance differences in initial ability for each of the two instructional groups, are recorded along with each table.

Usage.—Inspection of Table I makes it apparent that there was no significant difference between the adjusted means for these two groups, since the methods variance was smaller than the interaction variance. The hypothesis that there is no difference between these two methods with regard to the effectiveness with which they teach English usage cannot be rejected on the basis of these data.

Capitalization.—Table II presents the results of the analysis of covariance of the class means with regard to capitalization. Table II reveals that there was no significant difference between the adjusted means for the two groups involved. The hypothesis that there is no difference between the two methods employed, with respect to their effect on capitalization skills learned, is tenable on the basis of these data.

Punctuation.—Table III presents the analysis of covariance for these class means. The F of 1.74 computed from the methods and

Table I

The Analysis of Covariance of Initial and Criterion Class Mean Scores on Test 1—Usage

		SUMS OF SQUARES AND PRODUCTS			ADJUSTED VALUES		
Source of Variation	d. f.	ΣX^2	ΣXY	ΣY^2	d. f.	ΣY^2	Variance
M	1	5.06	4.59	3.89	1	.28	.28
M x S	18	1604.08	1489.05	1670.07	17	287.81	16.91

No value for F was computed, since the methods variance is smaller than the interaction variance.

Final means adjusted to allow for chance differences in initial ability.

Control groups : 39.27
Experimental groups : 39.26

Table II

The Analysis of Covariance of Initial and Criterion Class Mean Scores on Test 2—Capitalization

		SUMS OF SQUARES AND PRODUCTS			ADJUSTED VALUES		
Source of Variation	d. f.	ΣX^2	ΣXY	ΣY^2	d. f.	ΣY^2	Variance
M	1	15.28	69.29	336.24	1	150.68	150.68
M x S	18	1192.58	1925.53	7758.15	17	4649.71	273.51

No value for F was computed, since the methods variance is smaller than the interaction variance.

Final means adjusted to allow for chance differences in initial ability.

Control groups : 42.36
Experimental groups : 43.16

TABLE III

THE ANALYSIS OF COVARIANCE OF INITIAL AND CRITERION CLASS MEAN SCORES ON TEST 3—PUNCTUATION

		SUMS OF SQUARES AND PRODUCTS			ADJUSTED VALUES		
Source of Variation	d. f.	ΣX^2	ΣXY	ΣY^2	d. f.	ΣY^2	Variance
M	1	840.29	11.78	.16	1	440.62	440.62
M x S	17	4144.41	3349.43	6762.24	16	4055.30	253.46

F = 440.62/253.46 = 1.74
F required for significance at 5% level = 4.49

Final means adjusted to allow for chance differences in initial ability.
Control groups : 52.09
Experimental groups : 50.52

TABLE IV

THE ANALYSIS OF COVARIANCE OF INITIAL AND CRITERION CLASS MEAN SCORES ON TEST 4—GRAMMAR INFORMATION

		SUMS OF SQUARES AND PRODUCTS			ADJUSTED VALUES		
Source of Variation	d. f.	ΣX^2	ΣXY	ΣY^2	d. f.	ΣY^2	Variance
M	1	160.49	14.24	1.27	1	161.18	161.18
M x S	18	6066.58	6697.59	9303.05	17	1908.82	112.28

F = 161.18/112.28 = 1.44
F required for significance at 5% level = 4.45

Final means adjusted to allow for chance differences in initial ability.
Control groups : 50.46
Experimental groups : 49.71

TABLE V

THE ANALYSIS OF COVARIANCE OF INITIAL AND CRITERION CLASS MEAN SCORES ON TEST 5—SENTENCE STRUCTURE

		SUMS OF SQUARES AND PRODUCTS			ADJUSTED VALUES		
Source of Variation	d. f.	ΣX^2	ΣXY	ΣY^2	d. f.	ΣY^2	Variance
M	1	208.90	39.29	7.38	1	100.76	100.76
M x S	18	9838.39	8755.78	9606.28	17	1813.99	106.71

No value for F was computed, since the methods variance is smaller than the interaction variance.

Final means adjusted to allow for chance differences in initial ability.
Control groups : 55.43
Experimental groups : 54.75

interaction variance is too small for significance at either the one or five per cent level. The hypothesis that there is no difference between the two methods employed, with respect to their effect on punctuation skills learned, cannot be rejected on the basis of these data.

Grammar Information.—The analysis of covariance (Table IV) for these means shows a ratio (F = 1.44) between the methods and interaction variance too small for significance at either the one or five per cent level. On the basis of these data the hypothesis that there is no difference between the two methods with respect to their effectiveness in teaching grammar information is tenable.

Sentence Structure.—Table V shows the results of the analysis of covariance of the class means with regard to sentence structure. Since the methods variance is smaller than the interaction variance, it is apparent that there is no significant difference between the adjusted means of the two groups included in this analysis. Therefore, the hypothesis that there is no difference between the two methods employed, with respect to their effective-

ness in teaching sentence structure, cannot be rejected on the basis of these data.

Limitations

There were certain limitations that are apparently insurmountable in a study of this type. (1) It seems unlikely that all factors, other than those inherent in the methods themselves, which might operate systematically in favor of one method were completely controlled. The important part played by the teacher in an experiment of this type makes this variable especially important. Two procedures, namely, (a) the provision of specific lessons for the pupils and instructions to the teachers for their administration, and (b) the utilization of classes in each school taught by the same teacher, were employed in setting up this problem to control the teacher variable as closely as possible. It is apparent, however, that these procedures go only part way in removing the variable. It is quite conceivable that a teacher might have been more favorably disposed toward one method than toward the other, and consequently have varied in the enthusiasm and effectiveness with which she taught the two methods. (2) Interpretation of the results of this experiment must be made in terms of the evaluating devices used. No doubt, the tests used in this study were far from perfect and measured only a few of the many important elements of written composition. It is believed, however, that the element of composition most intimately related to diagraming, individual sentence organization, has been evaluated in a reasonably adequate manner. (3) Even though the diagraming test indicated that the experimental group as a whole had become quite proficient in diagraming sentences, it is quite possible that the instructional period was too short or otherwise inadequate to bring out the fullest possible value of this skill. The workbook from which the diagraming exercises were reproduced is intended to accompany composition writing and, therefore, to be used throughout a school year; however, the authors state that, if composition work is presented in one semester, this entire book should be completed in five school months. The book contains 245 pages. Judging from the authors' statement of time limits required, it seems logical to assume that the sixty-one pages and seven mastery tests that made up the lessons for the experimental group could be mastered adequately in eight weeks of intensive study.

Conclusions

In the light of the limitations stated above, the following conclusions seem defensible:

1. The learning of capitalization, punctuation, and English usage is no more pronounced under an instructional program composed largely of diagraming exercises than it is under a plan emphasizing composition exercises.

2. The diagraming of sentences is no more effective in teaching grammar information than is exercise on certain functional aspects of composition.

3. Sentence structure is taught as effectively by a composition method as by the diagraming of sentences.

Educational Implications

In the opinion of the writer, sentence diagraming is a mechanical skill that has little or no value in itself. That is, it is of no particular worth to a pupil to be able to analyze and graphically portray sentences over and above the improvement that it might bring about in his ability to write effectively. Furthermore, diagraming does not lend itself to correlation with any other subject or project in the program of the school.

In the light of these observations and of the above stated conclusions, there is considerable doubt concerning the advisability of employing sentence diagraming as a method of teaching language composition.

Bibliography

1. Barghahn, Kenneth C. "The Effects of Sentence-Diagraming on English Usage and Reading Comprehension." Unpublished Master's thesis, University of Iowa, 1940.
2. Barnett, Walter. "A Study of the Effects of Sentence Diagraming on English Correctness and Reading Comprehension." Unpublished Master's thesis, University of Iowa, 1941.
3. *Bibliography of Research Studies in Education*. Office of Education Bulletins (12 volumes). Washington: Government Printing Office, 1928–1940.

4. Cross, Ethan A., and Carney, Elizabeth. *Teaching English in High Schools.* New York: Macmillan Co., 1939.
5. Fries, Charles G. *American English Grammar.* New York: D. Appleton–Century Co., 1940.
6. Glaser, Emma. *On the Teaching of Junior High School English.* New York: D. Appleton–Century Co., 1935.
7. Greene, H. A. *Research in Elementary Language.* First Annual Research Bulletin, National Conference on Research in English. New York: Scott, Foresman and Co., 1933.
8. Greene, H. A. *Principles of Method in Elementary English Composition.* Fifth Annual Research Bulletin, National Conference on Research in English. New York: Scott, Foresman and Co., 1937.
9. Guiler, W. S., and Betts, E. A. *A Critical Summary of Selective Research in Elementary School Composition, Language, and Grammar.* Second Annual Research Bulletin, National Conference on Research in English. New York: Scott, Foresman and Co., 1934.
10. Knox, Opal. "Study of Sentence Structure as Related to Composition Quality." Unpublished Master's thesis, University of Iowa, 1941.
11. Leonard, J. Paul, and Arnell, Rebecca. "English-Language, Composition and Literature," *Review of Educational Research,* X (April, 1940), 107–125.
12. Lindquist, E. F. *Statistical Analysis in Educational Research.* Boston: Houghton Mifflin Co., 1940.
13. Lyman, R. L. *Summary of Investigations Relating to Grammar, Language, and Composition.* Chicago: University of Chicago Press, 1929.
14. Lyman, R. L. "Summary of Investigations Relating to Grammar, Language, and Composition (January, 1929, to January, 1931) I," *Elementary School Journal,* XXXII (December, 1931), 266–275.
15. Lyman, R. L. "Summary of Investigations Relating to Grammar, Language and Composition (January, 1929, to January, 1931) II," *Elementary School Journal,* XXXII (January, 1932), 352–363.
16. Lyman, R. L. "Summary of Investigations Relating to Grammar, Language and Composition (January, 1929, to January, 1931) III," *Elementary School Journal,* XXXII (February, 1932), 426–434.
17. Lyman, R. L. "Selected References in Elementary-School Instruction — English," *School Review,* XLI (February, 1933), 129–134.
18. Lyman, R. L. "Selected References on Elementary-School Instruction — Language," *Elementary School Journal,* XXXIV (October, 1933), 134–137.
19. Lyman, R. L. "Selected References in Elementary-School Instruction — Language," *Elementary School Journal,* XXXV (October, 1934), 135–138.
20. Lyman, R. L. "Selected References in Elementary-School Instruction — Language," *Elementary School Journal,* XXXVI (October, 1935), 125–128.
21. Parker, Roscoe E. *The Principles and Practice of Teaching English.* New York: Prentice–Hall, 1937.
22. Smith, Dora V. "The Contributions of Research to Teaching and Curriculum-Making in English (July, 1934, to July, 1937), *English Journal,* XXVII (April, 1938), 295–311.
23. Smith, Reed. "Grammar: The Swing of the Pendulum," *English Journal,* XXVII (October, 1938), 637–643.
24. Walker, Grace J.; Bartels, Nell F.; and Marye, Mary E. *Adventures Wise and Otherwise.* New York: Harcourt, Brace and Co., 1935.
25. Ward, C. H. *Grammar for Composition.* New York: Scott, Foresman and Co., 1933.

AN EVALUATION OF AN INDIVIDUAL METHOD AND A GROUP METHOD OF TEACHING COLLEGE FRESHMEN THE MECHANICS OF ENGLISH COMPOSITION

MARK KARP
State Teachers College
Paterson, New Jersey

THE METHODS TO BE EVALUATED

This experiment seeks to evaluate two methods of teaching college freshmen the mechanics of English composition. The students involved were divided into four groups: two of 27 each, one of 26, and one of 24. Because of illness and withdrawals, the classes finally numbered one of 27, one of 23, and two of 21. The group of 27 and one of the groups of 21 were designated as the Experimental Groups and the other two as the Control Groups. Each of two instructors taught one Control and one Experimental Section. The members of each of the Control Groups met three times a week with the entire group in attendance, for fifty minutes at each session. The Experimental Groups met three times a week for individual interviews. Five students, each for ten-minute interviews, were scheduled for each fifty-minute period. Two additional students were scheduled for each period in the event of absences. The Control Groups met for *group instruction*, and the Experimental Groups met for *individualized instruction*.

By the end of the semester, in January, 1941, the Control Groups had met for twenty-seven class sessions devoted to instruction in the mechanics of English composition. The time of the twenty-seven class sessions devoted to instruction in the mechanics of English composition was divided among the students of the Experimental Groups. By the end of the term each student of the Experimental Groups had met with his instructor seven times. Thus, the students of the Control Groups had met with their instructor for a maximum total of 27 x 50 minutes, or 1350 minutes; the students of the Experimental Groups had met with their instructor for a maximum total of 7 x 10 minutes, or 70 minutes.

For convenience in the discussion, the Experimental and Control Groups of one instructor will be referred to as E_1 and C_1, respectively, and the Experimental and Control Groups of the second instructor will be referred to as E_2 and C_2, respectively.

SOURCES OF THE DATA

All the students participating in the experiment took the following tests:

1. September, 1940: The American Council on Education, Psychological Examination, 1940 Edition.
2. September, 1940: Cooperative English Test, Form O.
3. January, 1941: Cooperative English Test, Form P.
4. April, 1941: Cooperative English Test, Form O.

The significance of the difference between the means of each Experimental Group and its corresponding Control Group on the intelligence examination is shown in Table I. There is a small difference of .76 in favor of E_1. The critical ratio of this difference is .30, which means that there are 62 chances out of 100 that the difference is a true difference.

In the second pair of groups there is a difference of .48 in favor of the C_2. The crit-

TABLE I

CRITICAL RATIOS OF THE DIFFERENCES OF MEANS OF THE EXPERIMENTAL AND CONTROL GROUPS ON THE AMERICAN COUNCIL ON EDUCATION PSYCHOLOGICAL EXAMINATION[1]

	GROUP			
	E_1*	C_1	E_2*	C_2
M	60.19	59.43	56.43	56.91
σ	7.61	8.78	9.50	5.29
σM	1.46	1.83	2.07	1.10
σD	.76		.48	
D	2.54		2.43	
CR	.30		.20	

* Subscript 1 indicates the groups of one instructor, and subscript 2 the groups of the second instructor.

ical ratio of this difference, .20, indicates that there are 58 chances out of 100 that the difference is a true difference.

In September, 1940, the Cooperative English Test, Form O, was administered to all the students. Table II presents the significance of the differences of the means of the Experimental and Control groups on this test.

TABLE II

CRITICAL RATIOS OF THE DIFFERENCES OF MEANS OF THE EXPERIMENTAL AND THE CONTROL GROUPS ON THE COOPERATIVE ENGLISH TEST, FORM O

	GROUP			
	E_1	C_1	E_2	C_2
M	57.82	56.13	52.91	53.71
σ	9.81	8.03	9.40	7.45
σ_M	1.89	1.67	2.05	1.63
D	1.69		.80	
σ_D	2.58		2.68	
CR	.65		.30	

It will be noted in Table II that in the first pair of groups there is a difference in favor of E_1 and in the second pair a difference in favor of C_2.

The critical ratio of the difference between the first pair of groups is .65. There are 74 chances out of 100 that the true difference is greater than zero, that is, there are 74 chances out of 100 that the M for the E_1 Group is greater than that for the C_1 Group.

The critical ratio of the difference of the means of the second pair of groups, .30, indicates that there are 62 chances out of 100 that the true difference is greater than zero.

It may be said, therefore, that on the basis of evidence in Tables I and II there was a difference in intelligence and achievement in the mechanics of English composition in favor of E_1 and a difference in these two areas in favor of C_2.

The Cooperative English Test, Form P, was administered in January, 1941, the end of the fall semester. Table III contains the significance of the differences between the Experimental and the Control Groups on that test.

An examination of Table III shows a difference in the first pair of groups in favor of E_1 and in the second pair a difference in favor of C_1. The critical ratio of the difference of the pair is 1.45, which means that there are 93 chances out of 100 that the difference is a true difference. The critical ratio of the difference of the second pair is 1.11, that is, there are 86 chances out of 100 that the difference is a true difference.

[1] The scores on all tests are presented in scaled score form, which means that there is direct comparability unit by unit among all scores. cf. John C. Flanagan, *The Cooperative Tests: A Bulletin Reporting the Basic Principles and Procedures Used in the Development of Their System of Scaled Scores.*

TABLE III

CRITICAL RATIOS OF THE DIFFERENCES OF THE MEANS OF THE EXPERIMENTAL AND CONTROL GROUPS ON THE COOPERATIVE ENGLISH TEST, FORM P

	GROUP			
	E_1	C_1	E_2	C_2
M	63.67	60.17	56.43	59.52
σ	9.07	7.63	8.50	9.12
σ_M	1.74	1.67	1.86	1.99
D	3.50		3.09	
σ_D	2.41		2.79	
CR	1.45		1.11	

In April, 1941, the Cooperative English Test, Form O, was readministered to test the retention of learning. Table IV contains the Critical Ratios of the means of the Experimental and Control groups on that test.

In Table IV it may be seen that in the first pair of groups a difference in favor of E_1 exists and that there is a difference between the second pair of groups in favor of C_2. The difference between the first pair has a critical ratio of 1.13. There are 87 chances out of 100 that this difference is a true difference. The critical ratio of the difference of the second pair of groups, .17, indicates that there are 56 chances out of 100 that the difference is a true difference.

TABLE IV

CRITICAL RATIOS OF THE DIFFERENCES OF THE MEANS OF THE EXPERIMENTAL AND CONTROL GROUPS ON THE COOPERATIVE ENGLISH TEST, FORM O, ADMINISTERED IN APRIL

	GROUP			
	E_1	C_1	E_2	C_2
M	65.04	61.91	57.62	58.10
σ	11.36	7.82	10.04	7.30
σ_M	2.18	1.63	2.19	1.59
D	3.13		.48	
σ_D	2.78		2.78	
CR	1.13		.17	

Tables V, VI, and VII show the gains from one test to another in each group and the significance of those gains from the first test to the second (Table V), from the second to the third (Table VI), and from the first to the third (Table VII).

TABLE V

CRITICAL RATIOS OF THE DIFFERENCES BETWEEN THE MEANS OF FORM O_s* AND FORM P OF THE EXPERIMENTAL AND CONTROL GROUPS ON THE COOPERATIVE ENGLISH TEST

Group	Mean of Form O_s	Mean of Form P	Difference of Means	CR of differences
E_1	57.82	63.67	5.85	5.70
C_1	56.13	60.17	4.04	5.69
E_2	52.91	56.43	3.52	3.16
C_2	53.71	59.52	5.81	6.38

* Subscript *s* designates the September administering of the test.

TABLE VI

CRITICAL RATIOS OF THE DIFFERENCES BETWEEN THE MEANS OF FORM P AND FORM O_a* OF THE EXPERIMENTAL AND CONTROL GROUPS ON THE COOPERATIVE ENGLISH TEST

Group	Means of Form P	Mean of Form O_a	Difference of Means	CR of Differences
E_1	63.67	65.04	1.37	.96
C_1	60.17	61.91	1.74	2.64
E_2	56.43	57.62	1.19	1.33
C_2	59.52	58.10	—1.42	1.25

* Subscript *a* refers to the April administering of Form O.

TABLE VII

CRITICAL RATIOS OF THE DIFFERENCES BETWEEN THE MEANS ON FORM O_s AND FORM O_a (COOPERATIVE ENGLISH TEST) OF THE EXPERIMENTAL AND CONTROL GROUPS

Group	Mean of Form O_s	Mean of Form O_a	Difference of Means	CR of Difference
E_1	57.82	65.04	7.22	6.36
C_1	56.13	61.91	5.78	8.46
E_2	52.91	57.62	4.71	5.71
C_2	53.71	58.10	4.39	5.05

Table V shows that all of the groups made significant gains in the knowledge of the mechanics of English Composition by the end of the teaching period.

Table VI presents the average gains and the significance of those gains made by each group between the test given at the termination of teaching in January and the test administered in April.

It is of note that all groups but one continued to gain in achievement after the teaching period had ended. The critical ratio of E_1, .96, shows that there are 83 chances out of 100 that the true difference is greater than zero. The critical ratio of C_1, 2.64, shows that there are 99.6 chances, almost certainty, that the true difference is greater than zero. The critical ratio of E_2, 1.33, indicates that there are 91 chances out of 100 that the true difference is greater than zero. The critical ratio of C_2, 1.25, indicates that there are 89 chances out of 100 that the difference is a true difference.

Table VII contains the significance of the differences between the mean scores on the test administered at the beginning and those on the test administered at the conclusion of the experiment. It will be seen in Table VII that all groups made significant gains. It can also be noted that each Experimental Group made greater numerical gains than its corresponding Control Group.

To ascertain which method is the better for superior students and which method is the better for below-average students, the six highest as a group and the lowest as a group of each Control and each Experimental section were matched with each other. Two patterns of matchings were made: one on the basis of intelligence, as revealed by the American Council on Education Psychological Examination, and the other on the basis of achievement in the mechanics of English Composition, as revealed by the Cooperative English Test, Form O_s, given at the beginning of the teaching period. The average scores on

TABLE VIII

EXPERIMENTAL AND CONTROL GROUPS OF INSTRUCTOR 1: GAINS IN ENGLISH OF THE SIX HIGHEST AND THE SIX LOWEST STUDENTS ON THE AMERICAN COUNCIL ON EDUCATION PSYCHOLOGICAL EXAMINATION

EXPERIMENTAL GROUP

ACE*		SCORES O_s** (Sept.)		P*** (Jan.)		O_a (April)		GAINS P—O_s (Sept. to Jan.)		O_a—P (Jan. to April)		O_a—O_s (Sept. to April)	
H	L	H	L	H	L	H	L	H	L	H	L	H	L
83	55	82	58	80	65	99	65	—2	7	19	0	17	7
74	54	66	46	65	56	75	55	—1	10	10	—1	9	9
68	53	70	65	75	65	87	67	5	0	12	2	17	2
67	50	43	52	49	60	51	57	6	8	2	—3	8	5
67	49	77	44	89	48	74	47	12	4	—15	—1	—3	3
65	48	55	52	61	52	64	55	6	0	3	3	9	3
						Total Gain		26	29	31	0	57	29
						Average Gain		4.33	4.83	5.17	0	9.5	4.83

CONTROL GROUP

74	51	60	52	66	54	59	52	6	2	—7	—2	—1	0
72	50	53	42	65	52	63	55	12	10	—2	3	10	13
70	49	59	66	65	65	65	65	6	—1	0	0	6	—1
69	48	51	53	53	57	59	63	2	4	6	6	8	10
68	45	69	51	67	59	78	62	—2	8	11	3	9	11
67	44	75	41	78	42	72	45	3	1	—6	3	—3	4
						Total Gain		27	24	2	13	29	37
						Average Gain		4.5	4	.33	2.17	4.83	6.17

* American Council on Education Psychological Examination
** Cooperative English Test Form O
*** Cooperative English Test Form P

these tests of these upper and lower groups were matched with each other. Because these samplings of the upper and lower ends of the distribution were small, it was felt that this method of matching instead of statistical treatment should be used.

Table VIII shows the gains of the six highest and the six lowest students of the Experimental and Control groups of one instructor on the psychological examination. It will be noted that in the Experimental section the lowest students show a slightly greater improvement than the highest at the end of the first period. Two of the highest showed very small losses, whereas two of the lowest showed no improvement. In the second period, from the termination of the teaching period to the conclusion of the experiment, the highest, except for one student, show a large improvement, as compared with the lowest, of whom three indicate slight losses and two no gain. An examination of the third pair of columns reveals that, except for one case in which a student showed a loss and another in which a student matched the achievement of a student in the lowest group, the highest students made greater gains than the lowest students. It would seem, therefore, that the Experimental method is better for the highest students than it is for the lowest. The implication of these data is that at least the most intelligent students of a group should have individualized instruction.

It can be seen in Table VIII that at the end of the first period the highest and the lowest of the Control section made approximately the same gain, with a slight difference in favor of the highest section. In the second period four students of the highest section showed either a loss or no gain as compared with two of the lowest section. The total gain shows a slight superiority for the lowest students.

A comparison in Table VIII of the total achievement of the highest of the Experimental Group with that of the highest of the Control Group shows a greater gain for the Experimental section. A further comparison in Table VIII of the achievements of the lowest sections of these two groups reveals gains for all students of the lowest section of the Experimental Group, a very slight loss

TABLE IX

EXPERIMENTAL AND CONTROL GROUPS OF INSTRUCTOR 2: GAINS IN ENGLISH OF THE SIX HIGHEST AND THE SIX LOWEST STUDENTS ON THE AMERICAN COUNCIL ON EDUCATION PSYCHOLOGICAL EXAMINATION

EXPERIMENTAL GROUP

ACE*		SCORES O_s** (Sept.)		P*** (Jan.)		O_a (April)		GAINS $P-O_s$ (Sept. to Jan.)		O_a-P (Jan. to April)		O_a-O_s (Sept. to April)	
H	L	H	L	H	L	H	L	H	L	H	L	H	L
76	53	62	43	73	46	74	43	11	3	1	—3	12	0
72	53	66	49	72	46	74	54	6	—3	2	8	8	5
65	51	53	56	61	56	61	61	8	0	0	5	8	5
63	42	57	48	55	53	64	50	—2	5	9	—3	7	2
62	41	61	40	64	43	60	42	3	3	—4	—1	—1	2
61	41	74	47	65	53	72	56	—9	6	7	3	—2	9
							Total Gain	17	14	15	9	32	23
							Average Gain	2.83	2.33	2.5	1.5	5.33	3.83

CONTROL GROUP

65	54	50	46	58	46	59	51	8	0	1	5	9	5
65	53	44	55	49	60	51	56	5	5	2	—4	7	1
61	53	50	56	49	66	55	57	—1	10	6	—9	5	1
60	49	60	42	65	53	62	52	5	11	—3	—1	2	10
59	49	52	55	64	65	67	60	12	10	3	—5	15	5
59	45	45	45	44	48	48	51	—1	3	4	3	3	6
							Total Gain	28	39	13	—11	41	28
							Average Gain	4.67	6.5	2.17	—1.83	6.83	4.67

* American Council on Education Psychological Examination
** Cooperative English Test Form O
*** Cooperative English Test Form P

and no gain for two students of the Control Group, and a small total difference in favor of the Control Group.

Table IX contains the gains in English of the six highest and the six lowest students of the Experimental and Control Groups of the second instructor on the psychological examination.

In Table IX a comparison of the achievement of the highest and the lowest students of the Experimental Group shows gains for four of the six in each group at the end of the first period and a small difference in favor of the highest group. During the second period four of the six in the highest group continued to show improvement as compared with three in the lowest group. Again there is a difference in favor of the highest group. The highest of this group made a larger total gain than the lowest, although four of the highest group showed a gain as compared with five of the lowest group.

In Table IX an examination of the achievement of the highest and lowest of the Control Group for the first period shows gains for four of the highest and gains for five of the lowest. Moreover, the lowest made a greater gain than the highest. In the second period five of the highest group showed improvement, whereas only one of the lowest group indicated improvement. The total gain of the highest students is larger than that of the lowest students of this group. These data suggest that at least the most intelligent students of a group should have individualized instruction in the mechanics of English composition.

The total gains of the highest of the Control Group, as revealed by Table IX, are greater than the total gains of the highest of the Experimental Group, and the lowest of the Control Group made a greater total gain than the lowest of the Experimental Group.

Table X is a presentation of the gains in English of the six highest and the six lowest of the Experimental and Control groups of one instructor on the initial achievement test, Form O of the Cooperative English Test, given in September.

Table X shows a gain in the first period for only two of the highest of the Experimental Group, as compared with gains for all six of

TABLE X

EXPERIMENTAL AND CONTROL GROUPS OF INSTRUCTOR 1: GAINS IN ENGLISH OF THE SIX HIGHEST AND THE SIX LOWEST STUDENTS ON THE COOPERATIVE ENGLISH TEST, FORM O

EXPERIMENTAL GROUP

SCORES						GAINS					
O_s (Sept.)		P (Jan.)		O_a (April)		P—O_s (Sept. to Jan.)		O_a—P (Jan. to April)		O_a—O_s (Sept. to April)	
H	L	H	L	H	L	H	L	H	L	H	L
82	49	80	55	99	48	—2	6	19	—7	17	—1
77	47	89	59	74	59	12	12	—15	0	—3	12
70	46	75	56	87	55	5	10	12	—1	17	9
70	44	65	48	68	47	—5	4	3	—1	—2	3
66	43	65	49	75	51	—1	6	10	2	9	8
66	37	65	48	69	44	—1	11	4	—4	3	7
					Total Gain	8	49	33	—11	41	38
					Average Gain	1.33	8.16	5.5	—1.83	6.83	6.33

CONTROL GROUP

75	52	78	59	72	62	3	7	—6	3	—3	10
69	51	67	53	78	59	—2	2	11	6	9	8
66	51	65	57	65	60	—1	6	0	3	—1	9
62	47	64	50	71	50	2	3	7	0	9	3
61	42	69	52	67	55	8	10	—2	3	6	13
61	41	63	42	67	45	2	1	4	3	6	4
					Total Gain	12	29	14	18	26	47
					Average Gain	2	4.83	2.33	3	4.33	7.83

the lowest of that group. During the second period five of the highest section made gains, as compared with only one of the lowest section. The total-gains column reveals gains for four of the highest and five of the lowest, with a small total difference in favor of the highest group.

Also in Table X it can be seen that four of the six highest in the Control Group showed gains during the first period, whereas all six of the lowest indicated gains. During the second period three of the highest showed gains compared with five of the lowest. The totals column indicates gains for four of the highest, gains for all six of the lowest, and a marked difference in favor of the lowest group.

In Table X it is seen that the highest of the Experimental Group made greater total gains than the highest of the Control Group and that the lowest in the Control Group made greater gains than the lowest of the Experimental Group. These data suggest that at least those who rank highest on a test in achievement in the mechanics of English composition, administered at the beginning of the teaching period, should have individualized instruction.

Table XI reveals that during the first period all the students of the lowest section of the Experimental Group made gains, whereas only two of the highest section made gains during this period. During the second period all of the highest made gains, as compared with only one of the lowest. The totals columns show a difference in favor of the highest section.

Moreover, in Table XI it can be noted that, although the highest of the Control Group made greater initial gains than the lowest, five of the highest during the second period showed losses, whereas five of the lowest showed gains. The totals columns indicate a difference in favor of the lowest.

If the total of the highest of the Experimental Group is compared with the total of the highest of the Control Group, it can be seen that there is a difference in favor of the Experimental section. A comparison of the lowest of the Control Group with the lowest of the Experimental Group reveals a difference in favor of the Control section.

The implications of the data in Tables X and XI seem to be: (1) that at least those who rank highest on a test in the mechanics of English composition administered at the beginning of the teaching period should receive individualized instruction and (2) that those who rank lowest on that test should receive group instruction.

TABLE XI

EXPERIMENTAL AND CONTROL GROUPS OF INSTRUCTOR 2: GAINS IN ENGLISH OF THE SIX HIGHEST AND THE SIX LOWEST STUDENTS ON THE COOPERATIVE ENGLISH TEST, FORM O

EXPERIMENTAL GROUP

SCORES							GAINS					
O. (Sept.)		P (Jan.)		O. (April)			P—O. (Sept. to Jan.)		O.—P (Jan. to April)		O.—O. (Sept. to April)	
H	L	H	L	H	L		H	L	H	L	H	L
74	48	65	53	72	50		—9	3	7	—3	—2	2
66	47	72	53	74	56		6	6	2	3	8	9
66	43	65	46	69	43		—1	3	4	—3	3	0
65	41	65	47	67	44		0	6	2	—3	2	3
62	40	73	47	74	47		11	7	1	0	12	7
57	40	55	43	64	42		—2	3	9	—1	7	2
						Total Gain	—2	30	25	—7	30	23
						Average Gain	.83	5	4.17	—1.17	5	3.83

CONTROL GROUP

71	50	82	49	74	55		11	—1	—8	6	3	5
68	46	68	46	75	51		0	0	7	5	7	5
60	45	65	44	62	48		5	—1	—3	4	2	3
59	45	66	48	65	51		7	3	—1	3	6	6
58	44	64	49	59	51		6	5	—5	2	1	7
58	42	67	53	59	52		9	11	—6	—1	1	10
						Total Gain	38	17	—18	19	20	36
						Average Gain	6.33	2.83	—3	3.17	3.33	6

CONCLUSIONS

On the basis of these data, eleven conclusions have been deduced:

1. At the beginning of the experiment there was a small difference in intelligence in favor of E_1 and a small difference in intelligence in favor of C_2 (Table I).

2. At the beginning of the experiment there was a small difference in achievement in English in favor of E_1 and a small difference in achievement in English in favor of C_2 (Table II).

3. At the conclusion of the teaching period there was a difference in achievement in English in favor of E_1 and a difference in favor of C_2 (Table III).

4. At the termination of the experiment there was a difference in achievement in English in favor of E_1 and a small difference in favor of C_2 (Table IV).

5. All groups made reliable gains from the beginning of the experiment to the end of the teaching period (Table V).

6. All groups but C_2 continued to make gains from the conclusion of the teaching period to the end of the experiment (Table VI).

7. All groups made reliable gains from the beginning to the end of the experiment (Table VII).

8. Under the Experimental procedure the more intelligent students made from the beginning to the end of the experiment greater average gains than the less intelligent students (Tables VIII and IX).

9. The less intelligent students under the Control procedure made from the beginning to the end of the experiment greater average gains than the less intelligent students under the Experimental procedure (Tables VIII and IX).

10. Rated on the basis of achievement in English before the experiment began, the highest students in the Experimental groups made from the beginning to the end of the experiment greater average gains than the lowest students in these groups (Tables X and XI).

11. Rated on the basis of achievement in English before the experiment began, the lowest students in the Control groups made from the beginning to the end of the experiment greater average gains than the lowest students in the Experimental groups (Tables X and XI).

THE NATURE AND CAUSES OF REGRESSIVE MOVEMENTS IN READING*

EVALYN BAYLE
Editor, Webster Publishing Co.
St. Louis, Mo.

From the time that Laval in 1879 called attention to the fact that the passage of the eye across a line of print is discontinuous, consisting of movement and pauses, psychologists have found the study of eye movements a provocative field. These movements are concerned with both a sensory process, perception, and a thought process, assimilation of ideas. Consequently they offer a method of examining the activities of the mind through a motor manifestation which is automatic and reflex in character.

Eye-movement studies have concerned themselves with patterns of eye movements exhibited for the various levels of learning, for different types and purposes of reading, and for oral as contrasted to silent reading. Attention has been called to possibilities for modifying patterns of eye movements through various types of training and teaching. In fact, the whole nature of the reading process has been clarified in the psychological laboratory through eye-movement studies.

As a result the terminology of eye movements has been standardized: the number of fixations in relation to the length of the line indicates the span of recognition; the duration of these fixations shows the speed of recognition; the number of regressive movements gives evidence as to the regularity of the reader's progress across the lines of print. Factors which affect each of these measures of the reading process have been studied and important conclusions reached.

Regressive movements are movements of the eyes backward rather than forward. The problem of regressions is important, because inherent in them is evidence of factors which interrupt the normal progress of reading. Regressions are regarded as unconscious adjustments of the eyes to some failure in the process of comprehension or of perception. Where the material read offers no real obstacle to perception or to comprehension, the eyes move forward in rhythmic fashion; when difficulties appear, their movements are interrupted as attention is diverted to a reconsideration of textual elements.

The fact that when the difficulty of textual material is increased, the number of regressions increases, both for mature and immature readers, has been demonstrated experimentally. The literature on eye movements has also considered what factors in general can be expected to produce regressions: difficulties in vocabulary or in thought, level of reading ability, purposes of reading, attitudes when reading, and the nature of the material read. Some analysis of the nature of regressions has been made in various studies: distinctions have been drawn between regressions at the beginning of a line caused by an inaccurate return sweep from the end of the preceding line, regressions caused by inability to attach meaning to a word, and regressions produced by mature readers who have made too long a forward movement and must regress to pick up what was missed. The possibility of reducing undesirable regressions has been demonstrated experimentally.

THE PROBLEM IN THIS STUDY

So far as the writer knows, no attempt has been made to study the causes of regressions within the line beyond calling attention to "difficulty." It seems probable that difficulties in interpreting meaning beyond those implied in the recognition and comprehension *per se* of words and phrases might, in thoughtful reading, produce regressions even for readers of mature habits. If similar difficulties in interpretation tend to cause the regressions to fall into patterns which resemble each other, this should provide evidence of value in understanding the mental processes present in apprehending meaning in reading. It is conceivable that regressive movements may represent the only possible method of adjustment to certain difficulties, and may

*A part of a dissertation submitted to the faculty of the Division of the Social Sciences in Candidacy for the Degree of Doctor of Philosophy, Department of Education, University of Chicago, 1941.

serve as evidence under certain conditions that the mind of the reader is engaged actively with the process of interpreting the material before him. Conversely, the absence of regressive movements for a reader who is faced with certain interpretative difficulties may indicate that he is reading mechanically, with little real understanding of the material his eyes are passing over.

The study reported here was undertaken to investigate the difficulties which cause regressions. It also included a study of the nature of regressions and an attempt to classify them according to pattern and cause.

The procedure was two-fold: (1) analysis of existing eye-movement records in order to discover the various patterns that regressions tend to fall into, and of the copy which produced these regressions to discover possible causes for the patterns; (2) setting up copy which incorporated the factors suspected of causing regressions, and testing out the hypotheses by making eye-movement records of skilled readers from this copy.

The preliminary study of regressive movements, based on eye-movement records of Selections 1 and 2 taken in the course of previous experimentation, resulted in several hypotheses as to the nature and the causes of regressive movements in reading. Regressions seemed to fall into six patterns: (1) initial regressions, (2) corrective regressions, (3) reference regressions, (4) word analysis patterns, (5) phrase analysis patterns, and (6) line analysis patterns. The causes for these patterns apparently lay in the nature of the text: difficulties in interpretation of meaning inherent (1) in word order, (2) in word groupings, (3) in a misleading juxtaposition of certain words, (4) in lack of punctuation to help make the meaning clear, (5) in shifts in the meanings of words, and (6) in the necessity for concentrating on key words or key elements in sentence units.

Also considered worth further study were these four things: (1) in the case of initial regressions, the manner in which the preceding line "broke" the thought, (2) the effect of the length of lines used, (3) the effect of the length of the selection being read, and (4) the effect of the "solidness" or "openness" of a line, as represented by the amount of white space used in it.

The first step, then, in the experimental study of regressive movements was to set up copy which would incorporate these factors in such a way that eye-movement records made while they were being read would test the hypotheses.

The process of preparing this copy was one of trial and error, various forms being tried out before the eye-movement camera until the final drafts were thought acceptable. In addition to including certain types of difficulty in the text to be used in the final experiment, many other problems had to be faced. It was decided to present the test material to the subjects in typewritten form. This necessitated writing the copy in such a way that the lines could be made fairly even in length; as a matter of fact, in three of the selections the right hand margin was kept even except for an occasional punctuation mark; in the fourth selection, lines were purposely made of different lengths. Second, it was necessary to locate the test difficulties within the lines in such a way that the effect of each could be kept separate from that of other factors to as great an extent as possible. Third, it was believed that the position of these difficulties should not be the same in each line, and an effort was made to distribute them among the first, second, and last thirds of the lines in which they appeared when they were of such a nature that they could be so distributed.

It was further assumed that the subjects for the experiment should have a chance to get into the trend of the selection and presumably "strike their stride" in reading before test factors which were to be studied should be introduced. Accordingly, an effort was made to write the selections in such a way that the reader's interest might be caught by the opening lines and carried on through the whole selection. The last lines only were counted as test lines, and those appearing before this (seven for Selection 3, and six for Selections 4 and 5) were considered as "warm-up" lines. Selection 1-x, for reasons to be explained later, consisted of only twelve lines in all.

A five-inch line was decided upon as being the best approximation of the standard reading page of typewritten material. However, because of the demands upon the copy already discussed and the necessity for having that copy present material which was interesting, certain variations were necessary. Selection 1-x, because it was an adaptation of

a selection used in the records examined in the preliminary study of regressions, has a line length of four and seven-eighths inches, and Selection 3 of the desired five inches. The nature of the copy used for Selection 5 produced lines varying from three to seven inches in length.

To give the subjects practice in reading before the eye-movement camera, it was decided to present all directions for the experiment in typewritten form, and to have them read before the actual photographing of eye movements should begin. Accordingly the directions were worked out so that their form appproximated that of the test selections, and were presented to the subjects in exactly the same way as the test materials. These directions were tried out in preliminary experimentation, as were the test selections, and wherever there was any ambiguity, changes were made to correct it. Graduate students in the Department of Education served as subjects in this preliminary testing.

Because limitations of the field of vision before the eye-movement camera prevent the use of too long a selection and because of the difficulty inherent in writing one continuous selection which should incorporate all the factors to be tested, three different test paragraphs were prepared. A fourth paragraph was adapted from Selection 1 as a kind of control selection. The copy used in the experiment was typed in pica type, double spaced, and was presented to the subjects without the titles used here for purposes of distinguishing the selections from one another.

The text of the materials used in the final experiment is given in the following pages. They are presented here in the same order as that in which they were placed before the subjects. Following the presentation of each passage significant characteristics will be discussed.

Selection 3 was intended to test the effect of shifts in the meanings of words upon eye-movement patterns. It also included certain other test factors.

This selection is an adaptation of a paragraph found in a story from the *Saturday Evening Post* of November 9, 1940. It was felt that its narrative style would capture the interest of the readers, and that the trend of its ideas would not be beyond their comprehension, even though its metaphorical style offered opportunity to study the effect of such phrases on the pattern of eye movements. The test lines begin with "stand extra expense," line 8. Enough of the slang appears before this point to enable the reader to adjust to it, so that the first test line should not produce a pattern of regressions due simply to a sudden introduction of something new and unexpected.

Two lines end with hyphenated words; it was expected that the next lines, beginning with the rest of the word in each case, might show an effect from this insofar as initial regressions are concerned. In one case, it is

SELECTION 3

Line

In August, 1919, the baby was born and her mother died. Tom came back to Chicago and brought the baby with him. He still had some dough in the bank, and he had this job with the *Register*. He had been with newspapers all his life and so he went back to work. It was up to him to learn to care for the baby also. He always had three squares a day, but he could not stand extra expense when he was down to buttons, on the shorts for groceries and rent and shoes for the baby. He became an expert at carefully fixing formulas, washing clothes, and bathing the small scoot. And of course she grew, grew like a bad wee. Finally came a day when she came paddling into where Tom was having his breakfast, swarmed up into his chair, and began to get off long flights of conversational noises, with gestures. She began to visit with him, constructing sentences of mouthings and blabberings which sounded like real conversation.

not hard to anticipate the ending of the word so hyphenated; in the other, the intention was to use a word in which the ending could not be so anticipated.

Scoot and *wee* are unusual, but not unintelligible, terms to substitute for *baby*. The use of *wee* in connection with *grew like a bad* might be expected to have one of two effects: either the reader would fail to notice that the expected word was not used, or he would be impelled to check up to be sure that he was not misreading. The use of an inversion (*came a day*) and a repetition (*grew, grew*) might have some effect on eye-movement pattern. Words like *paddling, swarmed, flights, mouthings,* and *blabberings* were introduced to test the effect of shifts in the meanings of words. To understand the trend of the passage, it is necessary for the reader to fit such words into the general context. This calls for the rejection of the basic meaning in favor of one which will fit this context.

The next to the last line was made to consist of words of more than average length, in contrast to lines made up of short words.

Selection 1-x was introduced as a kind of control selection. The text of this selection follows.

This selection is the same as Selection 1, used for the eye-movement records in a previous study which were examined in the preliminary study of regressive movements. Since, however, in this experiment the materials to be read were presented in typewritten form, it was necessary to make a few minor changes in the wording and punctuation in order to keep the lines even.

Since this selection has been used before and data are available as to the number of regressions produced in those studies, it was included in this study as a sort of control selection. Its simple form, vocabulary, and ideas made it seem probable that the present subjects might show their most normal pattern of eye movements in reading this selection, in contrast to the pattern which they were expected to show in reading the other test materials.

Selection 4 was planned to test the effect of unfamiliar words and of certain perceptual factors upon eye-movement patterns. The first test line in this selection, line 7, is made up of 11 short words, in contrast to line 17 of Selection 3 and line 14 in this selection, where there are only a few long words to the line. Line 8 introduces a vocabulary factor; it has been established that a common type of regression appears where the reader is under the necessity of trying to recognize an unfamiliar word, and this word *scriftgiser* is entirely unrecognizable, having been invented for the purposes of this experiment. This applies also to *gisscrifter* in the next to the last line. Line 9 was expected to produce a zigzag pattern across the whole line if the reader was involved in trying to make meaning out of the relation of the term *cowgirl* to *father*, as he should be if he is concerned with interpretation.

Two words are split with hyphens at the ends of lines to give additional evidence as to the effect upon initial regressions.

In three places misspellings are introduced; two of these consist in the doubling of a final letter at the end of the line. Such misspellings should be noticed by intelligent students if they fall well within the span of attention; however, since the last fixation on a line ordinarily is located a short distance in from the end of the line, it might be expected

SELECTION 1-X

Line

1 At the Battle of King's Mountain a young
2 Mr. Black was badly hurt. His colored servant
3 saw that his master had been left to die, and
4 carried him to the next town. There he hunted
5 until he found a doctor. The young man lived.
6 After the war, he gave the Negro a cottage on
7 his place. Always on the day of the Battle of
8 King's Mountain a big dinner was prepared for
9 the old servant. He sat alone at the head of
10 the master's table, and was served by the two
11 daughters of the family. This was done every
12 year that the old man lived.

SELECTION 4

Line

Books about the great northwest are very popular nowadays, and so are books which arise from the childhood days of the author. On the best-seller list for this month is the latest volume from the pen of Daniel Peterkin, entitled *Stranger Lives*. Memories of his own boyhood in the Canadian West are the source of ten of the stories in this book by a scriftgiser. Before he married and took up a homestead, the writer's father had been a cowgirl. He had been a hot-headed fist-shaker, finally settling down only after he began to farm. Father's picture in this book has warmth and a rich humann dignity. His bigest objection to farming was the gradually diminishing solitude when homesteaders settled nearby. This is handled by the skillfull gisscrifter with humor. "I just don't like having strangers in my pocket," father explained.

that these final letters would fall in the outer fringes of the perception span. It was felt that by incorporating such factors into the copy, some evidence might be obtained as to the need for corrective regressions. As a check upon this, one misspelled word, *bigest*, was placed within the line.

The last selection, the text of which is given next, presented certain other problems, chief of which were the effect of lack of punctuation and of lines varying in length.

The "warm-up" lines for this selection consist of a comment upon the material to be read, as a sort of introduction designed to arouse interest in and curiosity about the poem which follows. Since the lines of the poem are of uneven length, this preliminary material was set up with lines equally uneven

SELECTION 5

You are now to read a rather simple poem.
It is not written in regular form, with regular meter and rhythm,
but rather in a half prose style. This will make it
easier for you to read it. When you have finished the reading,
you should be able to tell what thoughts the poet had
in mind on each of the two occasions with which he deals here.

When I read a book, a biography famous,
And is this then said I what the author calls a man's life?
And so will someone when I am dead and gone write my life?
As if any man really knew aught of my life!
Why even I myself I often think know little or nothing of my
real life,
Only a few hints, a few faint clews and indirections
I seek for my own use to trace out in this book.

When I see a chest, a treasure chest,
And so, then, this, said I, is what men store their treasures in?
And thus will my treasures when I am dead and gone be stored away?
As if any man could really know my riches!
Really, I think, even I, the poet, do not know what my riches are,
Only some memories, faded clues to happiness
Which I hold precious for my own remembering.

and in about the same arrangement as the poem shows. This was for the purpose of accustoming the reader to the arrangement before he came to the actual test lines.

The first part of the poem is one of Walt Whitman's *Inscriptions*.[1] A few minor changes have been made in punctuation and in wording. Since it was not possible to expect comparable results from the lines of this selection and those in the other selections, because of the differences in length, a second stanza was prepared which would provide contrasts with the lines of the first one. In the first line of the first stanza, an inversion appears, but in the comparable line of the second stanza there is no such inversion of word order. Several lines in the first stanza are difficult to interpret because of lack of punctuation; in the second stanza lines comparable in difficulty of idea are punctuated. Care was taken, however, to vary the form of the line enough to obviate practice effect.

There are many places in this selection where the order of the words and phrases might cause difficulty in interpretation. Several vocabulary factors might also cause trouble: *aught, so, indirections, clews* (the spelling) were recognized as possible obstacles to interpretation. It will be noticed that the spelling *clues* is used in the second stanza. If the difference was noticed, it might produce a regression away from the next to the last line to the line in which the word was first used.

Selection 5 might be expected to produce the most initial regressions of the three test selections, if the length of line is a factor. Dearborn attributed initial regressions to the failure of the eyes to establish "short-lived motor habits" which enable a reader to repeat in subsequent lines that length of return movement which will enable him to begin the reading at the proper point. Obviously in reading lines varying so much in length no such habit is established.

The Experimental Procedures

The subjects used in the experimental study of regressions were ninth and tenth grade pupils from the University of Chicago High School. Thirty-five pupils from these grades were chosen on the basis of their percentile ranking on the Iowa Silent Reading Test. However, it was found later that 3 of these pupils reread so much on every selection that it was impossible to plot the films obtained from them, so that the size of the experimental group was cut to 32. Some records were lost during the experiment, but usable films were obtained from 31 subjects for Selection 1-x, 32 subjects for Selection 3, 31 subjects for Selection 4, and 28 subjects for Selection 5.

The median score on the Iowa Silent Reading Test for the tenth grade subjects was 104, which is equivalent to the ninety-seventh percentile. The median score for the ninth graders was 98, with a corresponding percentile rank of 96. No reader in either group was below the eighty-eighth percentile on the Iowa Silent Reading Test.

The median score for the tenth grade subjects on the American Council Psychological Test was 122, convertible into the eighty-seventh percentile. For the ninth graders, the median score was 114, which made their percentile rank on this test 89.

There were 13 tenth grade subjects and 19 ninth grade subjects. No attempt was made to keep the two groups separate. The whole group was considered as homogeneous insofar as reading and intelligence are concerned, and these factors only were considered as having a bearing upon the experimental results.

In testing out the hypotheses which were set up subjectively in the preliminary study as to the nature and causes of regressions, it was decided that only good readers should be used. It is known that people who approach mature habits of reading tend to make fewer regressions even in the presence of difficulties. Poor readers, on the other hand, tend to make many more regressions than good readers, even on easy material. It was expected that in the experiment regressions would be caused by test factors. The use of poor readers for subjects would have covered up the possible relationships between the difficulties introduced and their production of regressions, while the use of skilled readers tends to eliminate the types of regressions that are symptoms only of immaturity in reading skill.

The fact that there were appreciable variations between the various lines of the selections, both in number of fixations and number of regressions for the majority of the subjects, bears out this assumption. On easy and uncomplicated material, eye-movement patterns

[1] Walt Whitman, *Leaves of Grass*, p. 7. Emory Holloway, Editor. New York: Doubleday, Doran and Co., 1931.

characteristic of mature readers appeared; the pattern was altered when difficulties were introduced.

Comprehension tests on each selection were given following the taking of the eye-movement records. Five questions were asked on each selection. An effort was made to use questions which would test, not memory of detail, but the implications of each selection. It was decided beforehand that three correct answers for a selection would indicate adequate comprehension; four, very good comprehension; and five, excellent comprehension. By this measure, 6 per cent of the subjects did not have adequate comprehension on Selection 1-x and Selection 4, and 7 per cent on Selection 5. This means only two people in each case. On the other hand, 71 per cent had very good or excellent comprehension on Selection 1-x and Selection 4, 100 per cent on Selection 3, and 68 per cent on Selection 5.

Typical Eye-Movement Records on the Test Selections

Before analyzing the results from the experiment, it might be well to illustrate the kind of eye-movement patterns exhibited in the reading of the three test selections. Typical records for each are reproduced in Figures 1, 2, and 3. These records have been carefully chosen not only to represent the average number of fixations and regressions for the group as a whole, but also to show

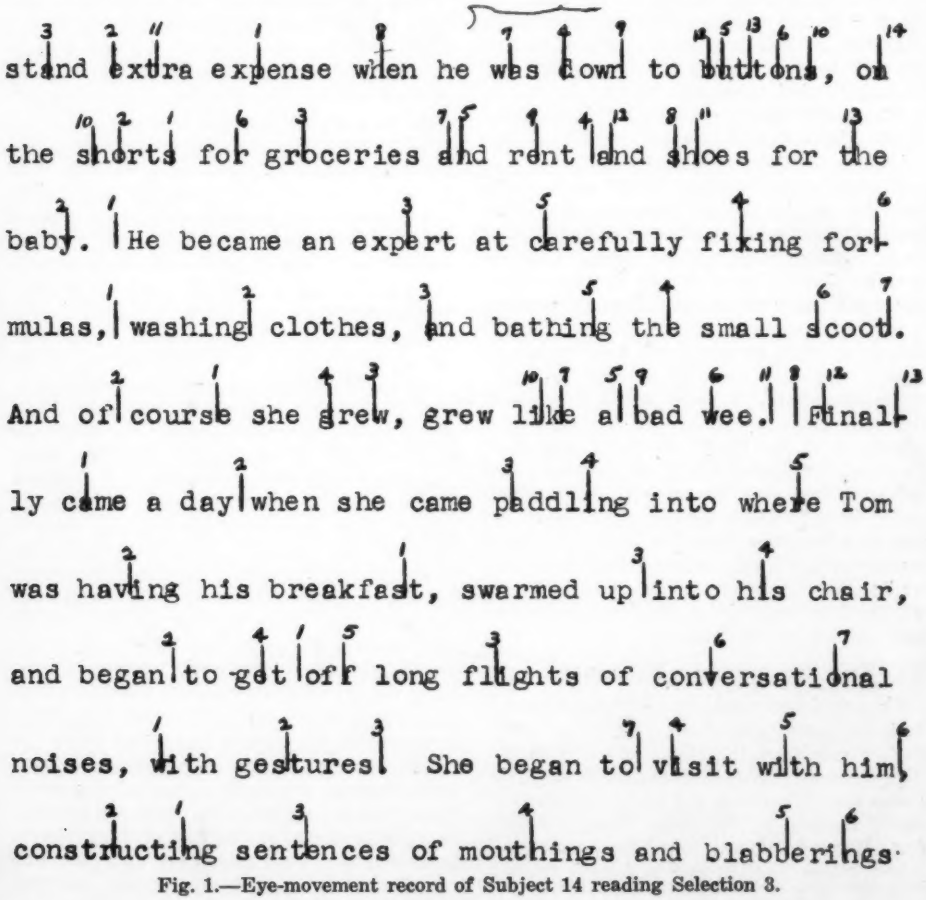
stand extra expense when he was down to buttons, on
the shorts for groceries and rent and shoes for the
baby. He became an expert at carefully fixing for-
mulas, washing clothes, and bathing the small scoot.
And of course she grew, grew like a bad wee. Finall-
ly came a day when she came paddling into where Tom
was having his breakfast, swarmed up into his chair,
and began to get off long flights of conversational
noises, with gestures. She began to visit with him,
constructing sentences of mouthings and blabberings.

Fig. 1.—Eye-movement record of Subject 14 reading Selection 3.

the typical positions that those fixations and regressions took over the test factors introduced into the copy. By following through the order of the fixations on each line, as shown by the numbering, the pattern of regressions can be seen. Later, typical lines will be graphed in detail to show this more clearly.

The most significant thing about these records is the wide variations that are revealed in the number of fixations and regressions for the different lines. Figure 1 shows the reading of Subject 14 for Selection 3. On Selection 1-x, the control selection, she averaged 6 fixations and 1.6 regressions per line. On Selection 3, the record of which is shown here, she varied from 4 to 14 fixations per line, and from none to 6 regressions, with an average of 8.2 fixations and 2.3 regressions. Such variations are typical of the majority of the subjects, and seem to be evidence in themselves of the effect of difficulties in interpretation. Where the line is easy and uncomplicated, the reader seems able to read with maximum efficiency; when difficulties appear, the eye movements reflect them in direct proportion to the amount of analysis they required before becoming clear.

Figure 2 shows how Subject 32 read Selection 4. Here again there are variations in number of fixations and regressions per line,

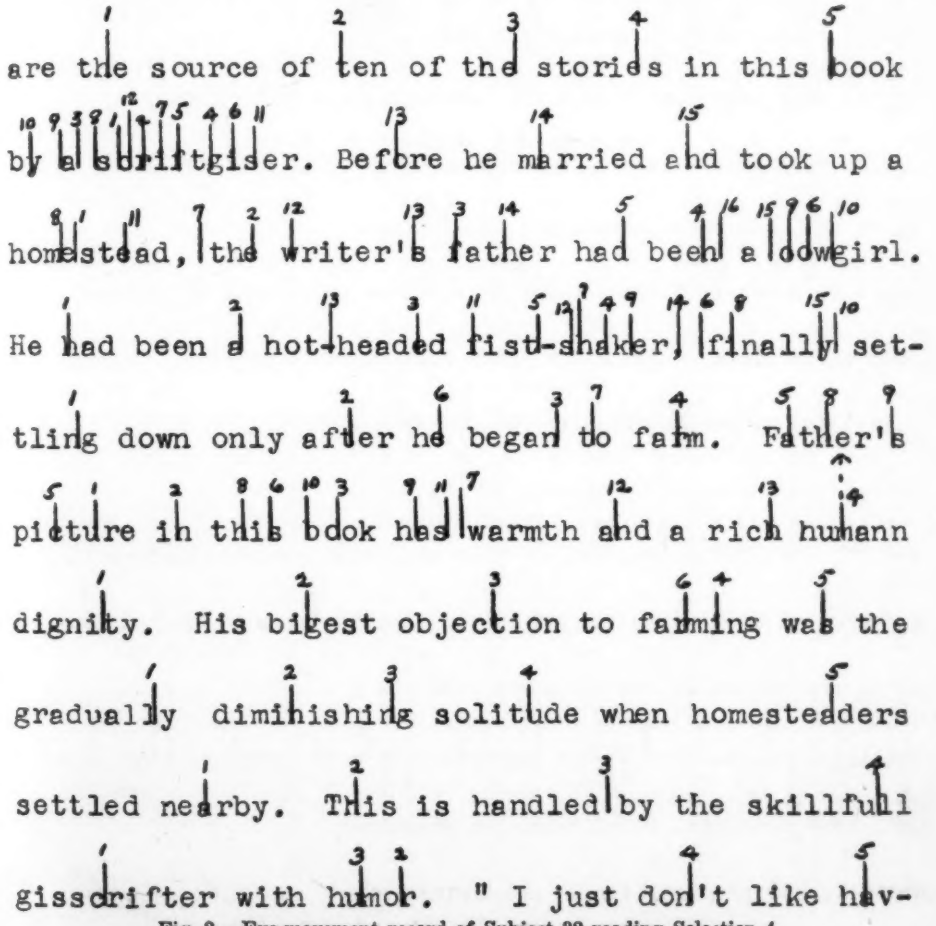
are the source of ten of the stories in this book
by a scriftgiser. Before he married and took up a
homestead, the writer's father had been a cowgirl.
He had been a hot-headed fist-shaker, finally set-
tling down only after he began to farm. Father's
picture in this book has warmth and a rich humann
dignity. His biggest objection to farming was the
gradually diminishing solitude when homesteaders
settled nearby. This is handled by the skillfull
gissdrifter with humor. " I just don't like hav-

Fig. 2.—Eye-movement record of Subject 32 reading Selection 4.

the fixations ranging from 4 to 16, and the regressions from none to 8, with averages of 9.3 and 2.4 respectively. On Selection 1-x this subject averaged 6.7 fixations per line and 1.3 regressions; on Selection 3, 6.5 fixations and 1.1 regressions. The record shown here for Selection 4 reveals the average for the group in number of fixations and regressions, and the pattern of eye movements again is illustrative of what happened when the majority of the subjects read this selection.

Figure 3 shows the reading of Selection 5 by Subject 12, with a total of 129 fixations and 31 regressions for the selection. Because of the difference in line length in this selection, the results cannot be compared with those on the other selections. However, two things stand out in this record, and in the records of practically every one of the 28 subjects from whom records were obtained on this selection. The first is the tendency of the reader to shorten the span of recognition and to move along regularly and cautiously over the lines. This may have been due to the effect of the "warm-up" lines which preceded the poem itself and which warned the reader of the questions he would be asked on the poem. Or it may have been due to the varying lengths of the lines. In any event, it is obvious that this selection was read cauti-

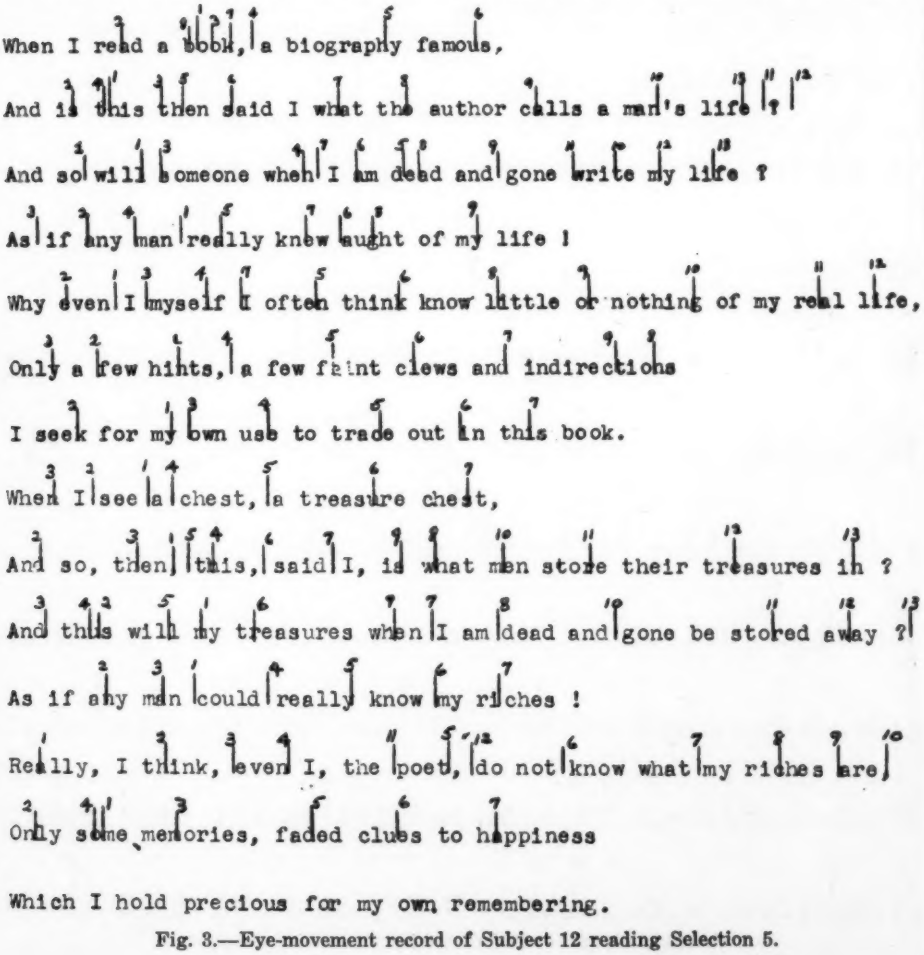
When I read a book, a biography famous,
And is this then said I what the author calls a man's life ?
And so will someone when I am dead and gone write my life ?
As if any man really knew aught of my life !
Why even I myself I often think know little or nothing of my real life,
Only a few hints, a few faint clews and indirections
I seek for my own use to trade out in this book.
When I see a chest, a treasure chest,
And so, then this, said I, is what man store their treasures in ?
And this with my treasures when I am dead and gone be stored away ?
As if any man could really know my riches !
Really, I think, even I, the poet, do not know what my riches are
Only some memories, faded clues to happiness
Which I hold precious for my own remembering.

Fig. 3.—Eye-movement record of Subject 12 reading Selection 5.

TABLE I

REGRESSIONS PRODUCED ON EACH LINE OF THE FOUR SELECTIONS

Line	Total Number of Regressions	Average Number of Regressions	Number of Initial Regressions	Per Cent of Subjects Making Initial Regressions	Number of Internal Regressions	Average Number of Internal Regressions	Per Cent of Total Number of Internal Regressions
Selection 1-x (31 Subjects)							
2	65	2.1	19	61	46	1.5	17
3	37	1.2	16	52	21	0.7	8
4	54	1.7	21	68	33	1.1	12
5	53	1.7	22	71	31	1.0	12
6	51	1.6	21	68	30	1.0	12
7	45	1.5	19	61	26	0.8	10
8	43	1.4	13	42	30	1.0	11
9	32	1.0	13	42	19	0.6	7
10	23	0.7	10	32	13	0.4	5
11	33	1.1	17	55	16	0.5	6
Total	436	14.0*	171	65	265	8.6*	100
Selection 3 (32 Subjects)							
8	83	2.6	22	69	61	1.9	14
9	65	2.0	17	53	48	1.5	11
10	47	1.5	17	53	30	0.9	7
11	67	2.1	25	78	42	1.3	9
12	95	3.0	25	78	70	2.2	16
13	58	1.8	16	50	42	1.3	9
14	61	1.9	19	59	42	1.3	9
15	70	2.2	16	50	54	1.7	12
16	52	1.6	18	56	34	1.1	8
17	40	1.3	18	56	22	0.7	5
Total	638	19.9*	193	60	445	13.9*	100
Selection 4 (31 Subjects)							
7	54	1.8	15	50	39	1.3	6
8	119	4.0	18	60	101	3.4	16
9	138	4.6	13	43	125	4.2	20
10	97	3.2	20	67	77	2.6	12
11	65	2.2	15	50	50	1.7	8
12	80	2.7	20	67	60	2.0	9
13	53	1.8	15	50	38	1.3	6
14	57	1.9	19	63	38	1.3	6
15	41	1.4	11	37	30	1.0	5
16	96	3.2	20	67	76	2.5	12
Total	800	26.7*	166	55	634	21.1*	100
Selection 5 (28 Subjects)							
7	82	2.9	25	89	57	2.0	8
8	122	4.4	18	64	104	3.7	14
9	109	3.9	22	79	87	3.1	12
10	94	3.4	27	96	67	2.4	9
11	77	2.8	19	68	56	2.0	7
12	67	2.4	25	89	41	1.5	5
13	58	2.1	23	79	35	1.3	5
14	62	2.2	25	89	37	1.3	5
15	68	2.4	16	57	52	1.9	7
16	92	3.3	20	71	72	2.6	10
17	65	2.3	22	79	43	1.5	5
18	81	2.9	17	61	64	2.3	9
19	59	2.1	25	89	33	1.2	4
Total	1,036	37.0*	284	78	752	26.9*	100

* Figured on the basis of the total number for the selection.

ously rather than with a free style, in which the reader forged rapidly ahead and made regressions when necessary.

The second thing that stands out in the records for Selection 5, and which is illustrated in the example given in Figure 3, is the difficulty the readers had in getting back to the beginning of the next line with any degree of accuracy when the preceding line was one of the longer ones. Only one line in this record for Subject 12 is without an initial regression, and six of the lines required two extra fixations in order to read the part of the line coming before the position of fixation 1. In many of the records on this selection, the process of returning to the beginning of the next line is facilitated by regressions part way back in the line just read.

It will be seen that there is a curious lack of points in Figure 3 at which fixations tend to pile up on words, or even on phrases, because of regression patterns. This is noticeable in the majority of the records on Selection 5.

Table I presents the complete data on regressions for the four selections used in the experiment. It will be seen that there are wide variations among the lines in the various selections in the number of regressions produced. The range in initial regressions, in per cent of the total number of subjects who made such regressions on the different lines, is from 32 to 96 per cent. The range in internal regressions, in terms of the per cent contributed to the total number for the selection by each line, is from 5 to 20 per cent. The wide range in each case can only be explained by the differences in the difficulty presented to the reader by the various lines.

Initial Regressions

There is no evidence from this experiment of any relationship between initial regressions and the position of the line in the selection or the sentence element beginning the line. There is ample evidence of a relationship between the distance of the first fixation from the beginning of the line and the need for an initial regression on that line. A tabulation made of the distances for all lines in Selections 3 and 4 is given in Table II. In measuring the distance of the first fixation in from the beginning of the line, all fixations within the first centimeter were put together, all those within the second, and so on. When the fixation came on the dividing line, it was counted in with the lower measure.

It will be seen from the data on a total of 600 lines given in Table II that the farther in from the end of the line the first fixation falls, the more likely it is to be followed by a regressive movement. Only 10 per cent of the fixations falling within the first centimeter of the line were followed by a regression, while each of those falling in the sixth and seventh centimeters was followed by a regression. There is a steady increase in the percentages as the distance of the first fixation from the beginning of the line increases.

This evidence of the relationship between the position of the first fixation on the line and the need for an initial regression leaves unanswered the question as to what causes the reader to fixate at an inappropriate distance from the beginning of the line. For the abnormal distances there is a possibility that the eye of the reader has been caught by some physical factor within the line, such as a punctuation mark or the configuration of a familiar word. There is the possibility also that such a first fixation may have fallen on the previous line, though there is no evidence in the films to show this. Ordinarily such a regression could be distinguished by the angle of the return sweep. Another possible explanation for such large inaccuracies in the return sweep is that the mind of the reader is occupied with the meaning of something just read and this first fixation represents a wandering of the eye rather than a fixation for perception.

The line beginning with a word continuation, *mulas*, in Selection 3 had 3 first fixations

Table II

Relationship of Position of the First Fixation to the Production of Initial Regressions

	Division of Line							
	1st Cm.	2nd Cm.	3rd Cm.	4th Cm.	5th Cm.	6th Cm.	7th Cm.	Total
Number of first fixations	83	235	193	67	16	5	1	600
Number followed by regression	18	125	144	57	15	5	1	355
Per cent followed by regression	10	53	75	85	94	100	100	59

falling within the first centimeter, 13 in the second, 10 in the third, and 5 in the fourth. The line in the same selection which produced the same percentage of initial regressions, 78, beginning with *And of course she,* had 2 first fixations in the first centimeter, 6 in the second, 15 in the third, 4 in the fourth, and one each in the fifth, sixth, and seventh centimeters. It would seem from this evidence that the *for-* at the end of the preceding line in the first case had caused readers to make a special effort to fixate accurately at the beginning of the line which finished the word. Even then, a regression was needed for the majority of the subjects to recognize the second part of the word. The line beginning with the last syllable of the word *finally* showed approximately the same number of first fixations in the second centimeter as for the line beginning with *mulas.* However, there was not the same need, apparently, in this line for a closer look at the word ending, and consequently there were 28 per cent less initial regressions for this line.

Perhaps there is some factor involved in the end of a line which in some cases impels the reader to an effort for accurate fixation at the beginning of the next line and in others impels him to take a chance on the success of a shorter return swing. In such cases the factor might be a physical one, such as a word continuation or a punctuation mark. It may be in other cases a matter of the interpretation of word combinations or of ideas. These interpretative factors probably will always be subject to individual variations. That the experimental records made with skilled readers showed variations, often large, between the effects of different lines on production of initial regressions seems to point to some such explanation.

Patterns of Regressions on Selection 3

Table III summarizes the data on patterns of internal regressions on Selection 3. It will be seen that the test factors caused difficulty for one-half or more of the readers for all test lines except lines 15 and 17. Fifteen per cent of the adaptations to difficulty were by shortened span rather than by regression patterns. Of the 110 cases where the adaptation was by regression patterns, 82 per cent were by phrase analysis and 18 per cent by line analysis.

Figure 4 shows the plotting of a typical phrase analysis pattern of regressions in the reading of line 12 in Selection 3. A separate printing of the line is used for each direction of movement of the eyes. The numbers above the lines representing the points of fixation indicate the order of the fixations, and the

Table III

Patterns of Regressions on Test Factors in Selection 3

Line	PATTERNS						Number of Cases[a]
	Word Analysis	Phrase Analysis	Line Analysis	All Patterns	Shortened Span	No Difficulty	
8 (buttons)	------	12	4	16	------	15	31
9 (shorts)	------	14	6	20	------	11	31
11 (scoot)	------	12	2	14	13	4	31
12 (wee)	------	18	3	21	5	5	31
13 (paddling)	------	11	1	12	6	13	31
14 (swarmed)	------	10	3	13	6	12	31
15 (get off)	------	12	1	13	------	18	31
17 (mouthings blabberings)	------	1	------	1	8	22	31
Total	------	90	20	110	38	100	248
Per Cent	------	82[b]	18[b]	45[c]	15[c]	40[c]	100

[a] One record could not be analyzed for pattern because of movement.
[b] Based on total number of patterns.
[c] Based on total number of cases.

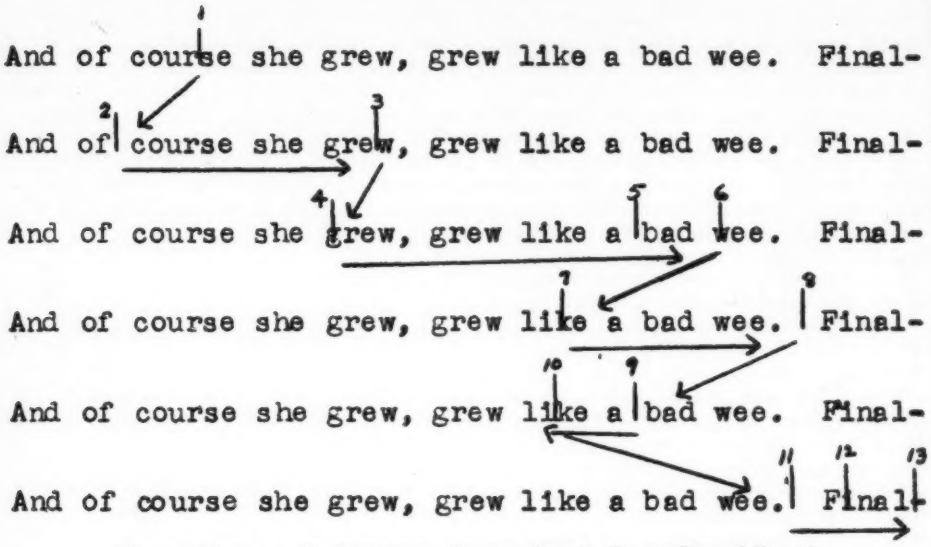
And of course she grew, grew like a bad wee. Final-
And of course she grew, grew like a bad wee. Final-
And of course she grew, grew like a bad wee. Final-
And of course she grew, grew like a bad wee. Final-
And of course she grew, grew like a bad wee. Final-
And of course she grew, grew like a bad wee. Final-

Fig. 4.—Phrase analysis pattern of regressions in the reading of line 12, Selection 3, by Subject 14.

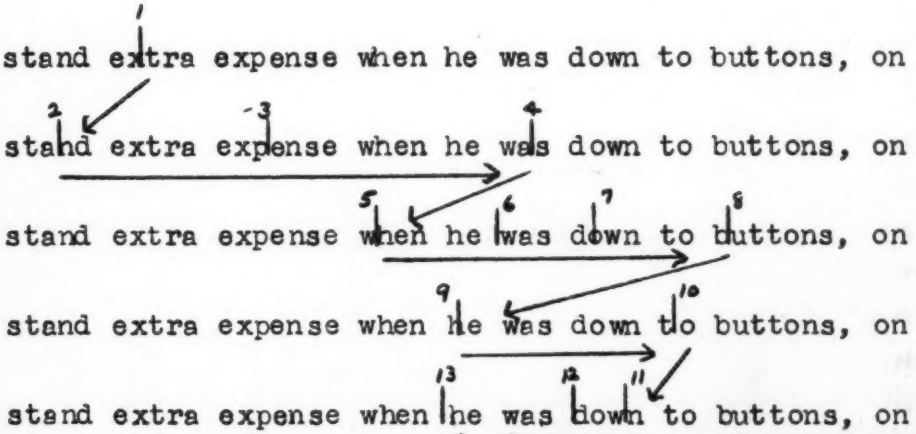
stand extra expense when he was down to buttons, on
stand extra expense when he was down to buttons, on
stand extra expense when he was down to buttons, on
stand extra expense when he was down to buttons, on
stand extra expense when he was down to buttons, on
stand extra expense when he was down to buttons, on

Fig. 5.—Phrase analysis pattern of regressions in the reading of line 8, Selection 3, by Subject 26.

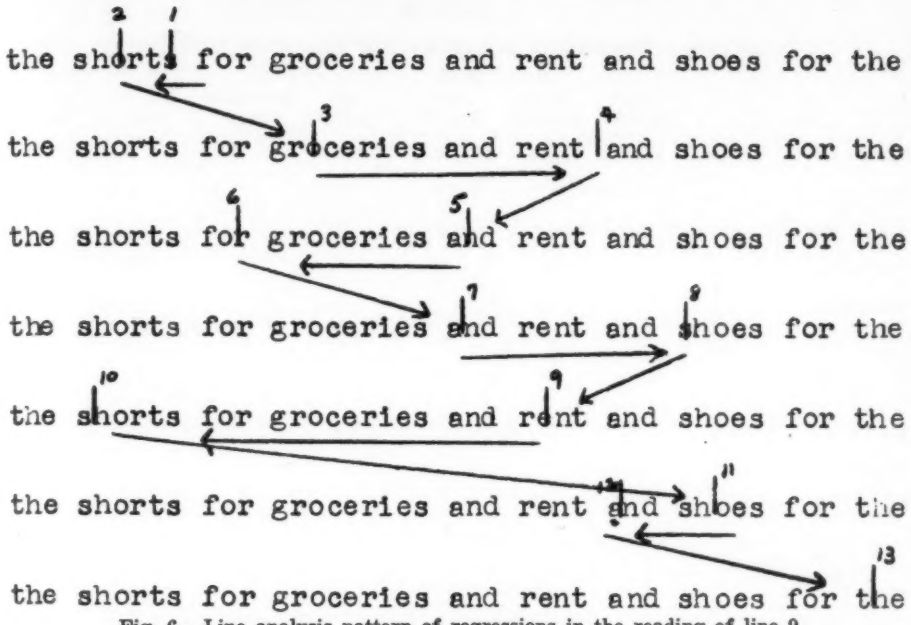
the shorts for groceries and rent and shoes for the
the shorts for groceries and rent and shoes for the
the shorts for groceries and rent and shoes for the
the shorts for groceries and rent and shoes for the
the shorts for groceries and rent and shoes for the
the shorts for groceries and rent and shoes for the
the shorts for groceries and rent and shoes for the

Fig. 6.—Line analysis pattern of regressions in the reading of line 9, Selection 3, by Subject 14.

arrows show the direction in which the reading is proceeding at the moment.

It will be seen that fixation 1 in this record falls quite far in from the beginning of the line and is followed by a regression. Fixation 4 may be a corrective regression, made to compensate for too long a forward movement from fixation 2 to fixation 3. The alternating forward and backward movements of the eyes over the phrase *like a bad wee* seem to indicate quite clearly that it was hard for this reader to interpret the combination of words.

Figure 5 shows a phrase analysis pattern over a part of line 8 in this selection. Figure 6 shows a line analysis pattern over line 9.

The phrase analysis pattern was the commonest form of adjustment to the type of test factor used in Selection 3. It is significant that, in attempting to interpret the meaning of such words as *buttons, scoot,* and *wee,* the reader examined the context in which it appeared. These words represent a kind of interpretative difficulty in which the basic meaning of the word must be rejected in favor of a figurative meaning. Apparently the process of interpretation in such cases is by examining the word in relation to the other words with which it appears through a regression pattern. In only 18 per cent of the cases where regression patterns appeared was line analysis used. The highest number of line analyses, six, was used in reading line 9, where the word *shorts* appeared near the beginning of the line; the largest number of adaptations by a shortened span was for *scoot,* which came at the end of the line. Apparently the position of the word in the line has some effect upon the type of adaptation made in its interpretation.

Patterns of Regressions on Test Factors in Selection 4

Table IV summarizes the data on number of regression patterns for Selection 4 on test factors which were expected to produce such patterns. It will be seen from these data that the factors which were introduced to test the hypotheses as to the causes and nature of regressive movements have produced the expected results in a significant number of cases. Lines 8, 9, 10, and 12 were expected to cause regressions; not only are the total numbers of regressions for these lines signifi-

TABLE IV

PATTERNS OF REGRESSIONS ON CERTAIN TEST FACTORS IN SELECTION 4

Line	PATTERNS: Word Analysis	Phrase Analysis	Line Analysis	All Patterns	Shortened Span	No Difficulty	Number of Cases[a]
8 (scriftgiser)	25	------	4	29	0	1	30
9 (cowgirl)	------	9	11	20	6	4	30
10 (hot-headed)	5	9	5	19	4	7	30
12 (picture)	------	11	6	17	3	10	30
16 (gisscrifter)	22	2	------	24	------	6	30
Total	52	31	26	109	14	28	150
Per Cent	48[b]	29[b]	23[b]	73[c]	9[c]	18[c]	100
12 (humann)	2	1	------	3	10	17	30
13 (bigest)	6	3	------	9	3	18	30
15 (skillfull)	1	1	------	2	7	21	30

[a] One record could not be analyzed for pattern because of movement.
[b] Based on total number of patterns.
[c] Based on total number of cases.

cantly high, but so also is the total number of patterns produced.

There were, for the thirty subjects whose records were such that they could be analyzed by patterns, a total of 109 regressive patterns on the five lines. This is 73 per cent of the total cases where such patterns were possible. Only 18 per cent of the cases were unaffected by the difficulty, while in 9 per cent adjustment was made by shortening the span of attention.

Of the total number of regression patterns, 48 per cent were word analysis. Of the 52 cases of word analysis patterns, all but 5 were caused by the two words *scriftgiser* and *gisscrifter*. Only one subject showed no effect on her eye-movement pattern by the presence of the first of these words. Two other subjects were stopped only momentarily by *scriftgiser*. One of these made four fixations on the word, but no regressions; the other returned to the word after reading over the rest of the line. The median number of fixations falling on the first third of the line containing this word for the remaining subjects was 9. The line seems to have been read in two ways: either the subject concentrated on this word and then went on to the rest of the line, or he left the word momentarily and then returned to have another try at it after reading the rest of the line. Many records show that the reading of the rest of the line was badly upset by, apparently, the tension introduced by failure to recognize and make meaning from *scriftgiser*. Figure 7 shows a typical word analysis pattern over this word.

The word *gisscrifter* in line 16 also produced typical word analysis patterns. Six subjects paid no extra attention to it and read the line through with great facility. Twenty-two subjects analyzed the word, and only two went beyond the word in a phrase analysis pattern that would indicate that they were trying to relate it to its context.

Line 9 in Selection 4 proved by all measures to be the hardest to interpret. The reason for this is obvious; the introduction of a fallacy in meaning is comparable as far as interpretation of words in context is concerned to the use of an unrecognizable word in line 8. That such a factor should cause more trouble than a vocabulary difficulty is thought-provoking. Four subjects read this line through with the type of eye-movement pattern that is expected of a very good reader. Perhaps it is significant that every one of the four had showed in his reading of the preceding line attempts to analyze the meaning of

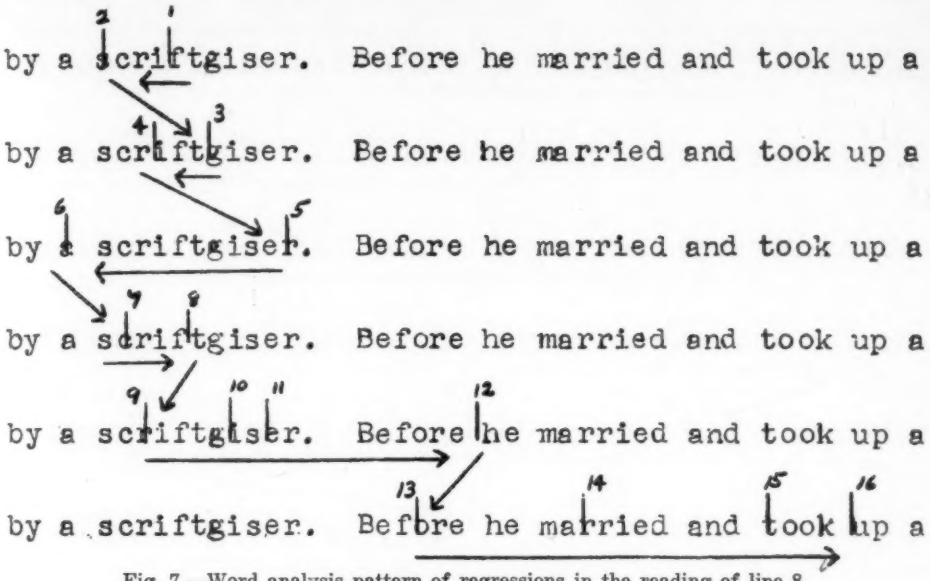
by a scriftgiser. Before he married and took up a
by a scriftgiser. Before he married and took up a
by a scriftgiser. Before he married and took up a
by a scriftgiser. Before he married and took up a
by a scriftgiser. Before he married and took up a
by a scriftgiser. Before he married and took up a

Fig. 7.—Word analysis pattern of regressions in the reading of line 8, Selection 4, by Subject 7.

scriftgiser. It would seem as though they recognized the similarity in the two situations and decided to react differently to this one. After they had read the test selections at the time of making the records, six subjects commented on the use of the word *cowgirl*. Three of these people were in this group who were unaffected, as far as their eye movements were concerned, by this factor in the line, so it would seem that they had taken in the word during their reading. Nineteen subjects mentioned *cowgirl* in the comprehension check.

There were three typical reactions to this line on the part of the rest of the subjects. Six of them read the last word carefully, but did not regress from it; 9 reexamined the second half of the line. Eleven reread the whole line, sometimes again and again. Figure 8 gives an example of analysis over the entire line.

The pattern of regressions for line 10 in Selection 4, which contributed 12 per cent to the total number of internal regressions for this selection, and which had an average of 2.6 regressions per subject on it, is for 9 subjects that of phrase analysis, extended here to cover the two words, ***hot-headed fist-shaker***. Seven subjects had no trouble with this line. Five subjects showed a line analysis pattern complicated enough to suggest great tension. Five people analyzed one word or the other, while four adjusted to the words by shortening their span of attention so that several fixations placed close together fell on the words.

Eleven subjects showed phrase analysis regression patterns for line 12 which indicated trouble in interpreting the first part of the line, ***picture in this book***. For only six was there analysis of the line. The predominance of phrase analysis as the favored reaction to this test factor is in keeping with the results on Selection 3, where it was clear that interpretation of shifts in the meanings of words calls for re-examination of the context in which the words appear.

The misspelled words at the ends of two lines in Selection 4 caused only 5 regression patterns in all, though there were 17 adjustments by shortened span of attention. For the misspelled word appearing within the line, there were more regression patterns (9) than adjustments by shortened span (3). In all three cases of misspelled words the subjects were apparently reading for ideas and 18 were not stopped by the misspelling, if they noticed it. Probably the fact that the extra letter on two of the words appeared at

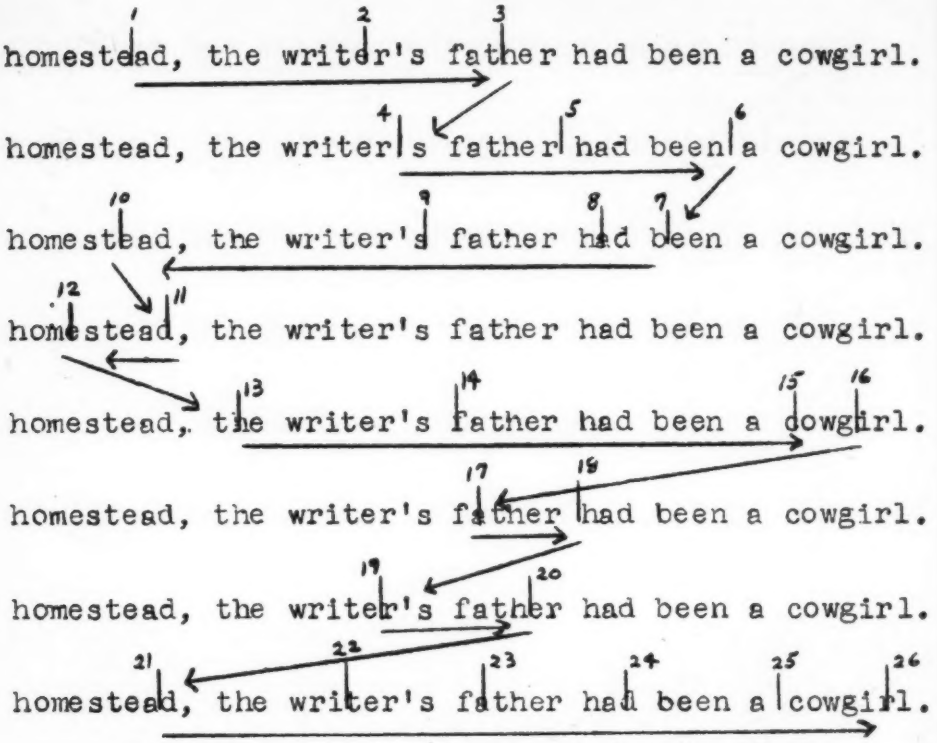
homestead, the writer's father had been a cowgirl.
homestead, the writer's father had been a cowgirl.
homestead, the writer's father had been a cowgirl.
homestead, the writer's father had been a cowgirl.
homestead, the writer's father had been a cowgirl.
homestead, the writer's father had been a cowgirl.
homestead, the writer's father had been a cowgirl.

Fig. 8.—Line analysis pattern of regressions in the reading of line 9, Selection 4, by Subject 9.

the very ends of the lines means that it was seen in each case with peripheral vision; if the reader's attention was caught by the misspelling, he looked again. This behavior may provide evidence as to what happens when corrective regressions are needed. Something on the fringes of peripheral vision to the right catches the reader's attention and makes him fixate too far to the right. Then he must make a regression to tie up the new perceptual unit with the preceding one.

One other factor remains to be discussed in relation to internal regressions in Selection 4. Two lines, 7 and 14, contrast sharply in the amount of white space they contain. Line 7 is made up of 11 short and very common words, while line 14 has only 5 words, which are easy to understand, but not quite so common as those in line 7. The results on these two lines also contrast sharply in the median number of fixations produced—9.1 fixations for line 7 and 6.2 fixations for line 14. The median number of internal regressions for line 7 is 1.4, and for line 14, 1.6 regressions. These figures seem to show that a line consisting of many short words is read with a larger number of fixations than is a line made up of a few long words. Apparently this factor of word length has little effect on regressions. The other evidence of this study points to the nature of the text rather than its physical appearance as a cause of regressions.

The last test line in Selection 3 consisted of only six words, *constructing sentences of mouthings and blabberings*. The medians for this line are 6.2 fixations and 0.1 regression. Selection 5 also offers evidence as to the effect of many short words.

PATTERNS OF REGRESSIONS ON SELECTION 5

As has been already pointed out, the patterns of eye movements on Selection 5 differ in many ways from those on the other test selections. There was a general tendency on the part of the subjects to adapt to difficulty by proceeding over the lines with a shortened span of attention and with fewer patterns produced by regressions. That 28 good readers all tended to read this selection in this way may be due to the abstractness of the ideas in the selection, to the varying lengths of the lines, or to the "openness" of those lines. Introspection suggests that the use of many short and familiar words in succession has a tendency to slow down our reading, because we cannot know what the key words are or group the words into familiar patterns. When such short and common words as those found in this selection are used in connection with longer words, we find it easier to find units of meaning among them. This is comparable with the reading of numerals, where attention must be paid to each digit, since the combination of digits is constantly varying and configuration is no help to recognition, as it is with words. In some such way as this the reading of Selection 5 may have been affected by the "openness" of the lines and the type of vocabulary used. The evidence from line 7 in Selection 4 suggests this kind of reaction to such lines. However, the need for grasping an abstract idea may have had its effect also, though it might be argued that this would lead to regression patterns rather than to a shortened span of attention. The reader must get an overview of what he is reading before he can adapt to ideas, while perceptual factors present themselves immediately. The tendency to use longer fixations for this selection than for the others was apparently an adaptation to the requirements of interpretation.[1]

The data on the regressions for each line of Selection 5 cannot be discussed in the same way as for Selections 3 and 4, because of the fact that the lines were not of the same length in Selection 5 and the totals are affected by this. But when the results for lines of similar length are compared, rather striking contrasts are noted. Table V presents the data on internal regressions for Selection 5 in such a way that these lines can be compared. It will be seen from this table that there are differences between lines of the same length and comparable idea, larger for lines 8 and 15 than for the other pairs. The advantage is with the line appearing in the second stanza of the poem in five cases and with the line appearing in the first stanza in the other case. Because of the difficulty of telling when the subjects stopped reading the last line, it cannot be plotted, so no comparison can be made between lines 13 and 20.

[1] The average duration per fixation for all subjects was 7.0 thirtieths of a second for Selection 1-x, 7.1 thirtieths for Selection 3, 7.3 thirtieths for Selection 4, and 7.4 thirtieths for Selection 5.

TABLE V

INTERNAL REGRESSIONS ON COMPARABLE LINES IN SELECTION 5

Line	Internal Regressions	Average For Line	Per Cent Contributed to Line
7	57	2.0	8
14	37	1.3	5
8	104	3.7	14
15	52	1.9	7
9	87	3.1	12
16	72	2.6	10
10	67	2.4	9
17	43	1.5	5
11	56	2.0	7
18	64	2.3	9
12	41	1.5	5
19	33	1.2	4

Line 7 is comparable to line 14 in length and structure, except for the inverted order in *biography famous*. Too much reliance cannot, however, be put upon the difference in the number of internal regressions on the two lines. Although six lines had been read before the subjects began the actual poem, the difference in the nature of the preliminary lines and those in the poem may have caused an unnatural manner of reading line 7. However, 12 subjects showed an analysis pattern resulting from regressions over the phrase, *biography famous*, which may have been caused by the inverted order or by the process of interpreting the words. No such pattern is shown in the records on line 14 for ten of these subjects.

Line 8 is comparable to line 15. There is a difference of 52 internal regressions in the total numbers for the two lines. The average per subject for line 8 is 3.7 internal regressions and for line 15, 1.9 internal regressions.

The variable factor in these lines is the punctuation introduced in line 15 in contrast to the absence of punctuation in line 8. The thought in the two lines is different enough to obviate practice effect upon the second from the reading of the first.

In only 3 records out of the 28 secured on this selection was line 8 read with the same facility as line 15. In 5 records the subjects seemed to have difficulty with both lines. The remaining 20 subjects showed by patterns of regressions on line 8 that they were finding the line hard to interpret, and showed fairly regular progress across line 15.

For 25 subjects who had difficulty with line 8 two patterns stand out. Fourteen subjects showed a tendency to analyze the first third of the line. Eleven subjects adapted to line 8 by analysis of the whole line. Figure 9 illustrates adaptation by line analysis. Subjective evidence tends to show that the lack of punctuation in line 8 makes it very hard to interpret. Without punctuation, the relationships of the words to each other must be puzzled out by rereading. Probably the subjects who showed phrase analysis patterns on the first part of the line were affected by the lack of punctuation. The other subjects who analyzed other parts of the line may also have been affected by the lack of punctuation, but were probably also affected by the need to answer a specific question on the idea expressed in the line, mentioned in the preliminary lines on this selection.

Examination in detail of the patterns of eye movement on the other lines shows no great differences in the patterns caused by regressions, with the exception of line 10. The rather sharp contrast seen between the numbers of internal regressions for lines 10 and 17 is probably due to a vocabulary factor in line 10. The word *aught* caused regressions indicating word analysis for 11 subjects, phrase analysis for 2 subjects, and line analysis for 2 subjects. No such difficulty is found in line 17.

With these exceptions there is great uniformity for each subject in the patterns of eye movements for the pairs of lines. The few places where fixations tend to pile up on a line, never outstanding as in Selections 3 and 4, seem to be determined entirely by individual reactions to the general implications of the ideas presented. Even for line 18, where the number of regressions rises over that for line 11, there is no uniform factor of difficulty.

Summary and Interpretation

Cause of regressions.—Analysis of the data from this experiment provides striking support for the following hypotheses concerning the causes of regressive movements in reading.

1. Regressions appear when some factor interrupts the flow of thought during reading, as well as when perception is inadequate.

2. Inadequate perception may be the result of eye movements which overreach or

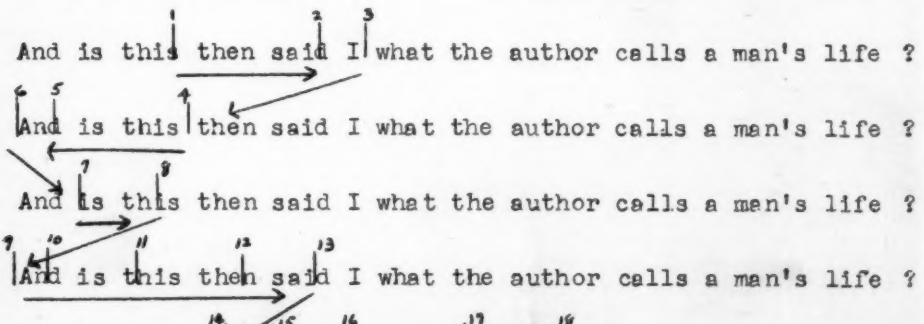
And is this then said I what the author calls a man's life ?
And is this then said I what the author calls a man's life ?
And is this then said I what the author calls a man's life ?
And is this then said I what the author calls a man's life ?
And is this then said I what the author calls a man's life ?
And is this then said I what the author calls a man's life ?

Fig. 9.—Line analysis pattern of regressions in the reading of line 8, Selection 5, by Subject 14.

underreach the limits of the reader's span of recognition.

3. Failure to recognize the basic meaning of a word may interrupt the continuity of thought processes during reading. An interruption of this nature will be characterized by regressions which analyze the word before the reading proceeds.

4. Failure to recognize the correct meaning for a word within its context may also cause regressions while an unsuitable meaning is being rejected in favor of one which fits into the context.

5. Failure to combine the meaning of a word properly with that of other words with which it appears may cause regressions.

6. Failure to relate the meaning of the word to the conditions under which it is used may interrupt the thought processes going on during reading, and this interruption will be reflected in regressive movements.

In addition to supporting the foregoing hypotheses, the data from the experiment suggest that failure to relate word meaning to context may be the result of a misleading connotation of the word. It may also be due to the fact that the referent of the word is not clear to the reader. Figurative language may also cause such failure to understand a word correctly in context.

Furthermore, the data seem to indicate that failure to combine the meaning of a word correctly with that of other words may be due to an inability to grasp sentence relationships, to a misunderstanding of word functions, or to a failure to catch a shift in the meaning of words.

When the reader is not alert to the author's mood and intention, he may not relate the meaning of a word to the conditions under which it is used. He may then be forced to regress in order to analyze apparent inconsistencies.

The analysis presented in this study reveals very clearly that the large majority of regressions serve the purpose of permitting the reader to re-examine all available clues to meaning within the material read. Consequently they represent a mode of adjustment to difficulty in interpretation. As such, they must be regarded as necessary parts of the reading process under those conditions where detailed search for meaning is required, in contrast to situations calling for simple assimilative reading.

The nature of regressions.—Difficulties in reading which call for analysis are reflected in the nature of the regressions used. This study supplies convincing evidence of the validity of the hypotheses to the effect that regressions tend to fall into six patterns. The first two of the patterns are related to inadequate perception, the last four to inadequate interpretation.

1. Initial regressions are adjustments following the first fixation on a line when the reader has failed to connect the end of the preceding line with the beginning of this one, either physically, with the proper arc of movement of the eyes, or mentally, with a suitable joining of the ideas represented.

2. Corrective regressions are internal regressions which serve to tie up fields of vision when the forward movement of the eyes in reading has overreached the span of recognition. The regression is used to make the necessary adjustment for successful perception, and resembles the initial regression in this respect.

3. Reference regressions are brief reexaminations of a textual element and serve the purpose of verification or of recall. They are infrequent and are, as the name implies, rapid references to something already read when a review of that element will assist in the interpretation of the new material under consideration.

4. Word analysis patterns of regressions are limited to a reexamination of a single word, and come about when the form or basic meaning of the word does not rise into consciousness during perception.

5. Phrase analysis patterns of regressions are adaptations to difficulty in apprehending the meaning of a combination of words, and serve to review several words in sequence. They are related to failure to recognize the meaning of a word within its context, failure to combine the meaning of a word with that of other words with which it appears, and failure to relate the meaning of a word to the conditions under which it was used. For skilled readers this is the most frequent pattern of regressions.

6. Line analysis patterns of regressions are characterized by a reexamination of the whole line. This pattern is a method for meeting failure to get the implications of a still larger unit of meaning than that which causes the phrase analysis pattern.

Any one of the three analysis patterns—word, phrase, or line—may arise from failure to interpret, as well as from attempts to interpret.

The first two patterns of regressions—initial and corrective—are likely to appear in all kinds of reading. They may be more frequent for skilled readers than for unskilled, inasmuch as such readers show a wider span of recognition than unskilled, and so are more likely to overreach in their forward movements. These patterns may also be more frequent in simple assimilative reading than in analytical reading, since in such reading the forward movement is likely to be more rapid and more confident.

The last four patterns of regressions are not likely to appear so often in simple assimilative reading as they are in reading that requires thoughtful interpretation, though they may appear even in easy material for readers whose habits of attack upon textual materials are immature.

Significance of study.—The findings of this study emphasize two facts: first, that regressions of the analysis pattern types serve a definite purpose in reading; and second, that the number of regressions needed for such patterns is dependent upon the success of the interpretation resulting from them. These have important implications for all teachers who are concerned with the teaching of reading or with guidance in reading.

Judd[1] has pointed out that the problem of "confusion periods" resulting from failure to recognize words must be met by adequate methods of teaching word analysis. Judd and Buswell[2] have pointed out that teachers must guide pupils in making intelligent attacks upon textual materials when "study" situations cause confusion in method and purpose.

Methodology in reading has supplied means for meeting the situation insofar as word recognition is concerned. However, as yet no adequate teaching methods have been devised for training pupils in meeting the problems of interpretation inherent in shifts in the meanings of words, changes in the author's mood and intention, figurative language, word relationships, and sentence structures. Research is needed to develop methods adapted to these situations.

Then, again, the whole field of the teaching of reading has suffered from failure to draw clear-cut distinctions between simple assimilative reading and that type which calls for careful interpretation. Research should be concentrated upon finding techniques for meeting both situations; so far it has been concerned primarily with the first type.

Recognition of the difference between these types of reading, and appreciation of the problems inherent in the second type, are the business of those teachers who develop reading programs, as well as of those who teach reading. It has been shown in this study that even skilled readers are not as efficient as they might be in meeting the sort of difficulty that much of their reading presents. Pupils must be helped in every way possible to meet the demands interpretation puts upon them. They must not be discouraged from regressing when regressions serve an honest purpose. They must be shown where to look for help within the text, and how to make use of reexamination of textual materials as an aid in successful interpretation. Only in this way may a serious danger be avoided—the tendency toward superficial reading brought about by too much emphasis upon techniques suitable only for simple assimilative reading situations.

[1] Charles H. Judd, *Reading: Its Nature and Development*. Supplementary Educational Monographs, No. 10. Chicago: University of Chicago Press, 1918.

[2] Charles H. Judd and G. T. Buswell, *Silent Reading: A Study of the Various Types*. Supplementary Educational Monographs, No. 23. Chicago: University of Chicago Press, 1922.

AN EXPERIMENTAL STUDY OF THE RELATIONSHIP OF PERIPHERAL PERCEPTION TO FACTORS IN READING[1]

CYRUS W. LAGRONE, JR.

PROBLEM

The purpose of this investigation was to determine the manner in which reading ability and reading habits are related to accuracy of perception in peripheral vision. More specifically, the purpose of the study was to find out how factors involved in or associated with reading, such as intelligence, spelling, vocabulary, reading speed, reading comprehension, reading span, and regressive movements, are related to: (1) accuracy of perception in the right peripheral field of vision, (2) accuracy of perception in the left peripheral field of vision, and (3) preference for, or comparative superiority in, the right peripheral field.

Many writers have defined the importance of peripheral perception in reading, and the way in which peripheral stimuli operate in the reading process. Dearborn points out that "reading, like efficient driving of an automobile in traffic, demands a very skillful use of peripheral vision."[2] Dodge[3] has explained that as the reader's eyes move across the line, the cues received through peripheral vision operate as partial perceptions which furnish premonitions of coming words and phrases. Furthermore, Dodge[4] and also Meumann, cited by Tinker,[5] contend that the word forms in peripheral vision act as stimuli for the reflex movement of the eyes across the line. Gray,[6] in his discussion of the factors involved in the anticipation of meaning in reading, emphasizes the importance of "the marginal impressions made upon the retina"; these, he contends, function in reading as "preperceptions which initiate the process of interpretation."[7]

An impressive array of experts in the field thus seem agreed in stressing the importance of peripheral factors in reading. Consistent with this emphasis, Dearborn interprets the reading disability caused by aniseikonia,[8] presumably a foveal anomaly, as seemingly attributable to its interference with peripheral vision. In his own words: "It is with this peripheral view of the line of print, which is so requisite for the proper spacing of the fixation pauses, that aniseikonia may interfere."[9] The question of optimal length of line, as Dearborn points out, is also related to peripheral factors. In this connection, he makes the observation that "those lines are best suited to rapid reading which give opportunity for a wide span of attention, but which are not of such length that the peripheral perceptions from the end or beginning of a line are too inexact and confused to be of value in determining the general character of a large part of the line."[10] It is thus generally accepted by students of the reading process that, without the perceptions in peripheral vision, reading would be both slow and difficult. Peripheral stimuli help govern the width of the reading span, as well as the duration of individual fixations.

Notwithstanding the importance that has been attributed to peripheral perception in reading, few data have been gathered concerning the manner in which individual variations in peripheral perception are related to reading. Although Ruediger[11] approached this problem in his study of the relationship between reading and the area of clear vision, except for the work of Crosland[12] there is little experimental evidence bearing directly on this problem.

[1] This article is a part of a Doctor's dissertation in educational psychology at the University of Texas.

[2] W. F. Dearborn, "On the Possible Relations of Visual Fatigue to Reading Disabilities," *School and Society*, LII (1940), 536.

[3] R. Dodge, "An Experimental Study of Visual Fixation," *Psychological Review, Monograph Supplements*, Vol. VIII, No. 4 (1907).

[4] *Ibid.*

[5] M. A. Tinker, "Visual Apprehension and Perception in Reading," *Psychological Bulletin*, XXVI (1929), 231.

[6] C. T. Gray, "The Anticipation of Meaning as a Factor in Reading Ability," *Elementary School Journal*, XXIII (1923), 614–626.

[7] *Ibid.*, p. 615.

[8] See W. F. Dearborn and I. H. Anderson, "Aniseikonia as Related to Disability in Reading," *The Journal of Experimental Psychology*, XXIII (1938), 559–577.

[9] W. F. Dearborn, "On the Possible Relations of Visual Fatigue to Reading Disabilities," *School and Society*, LII (1940), 536.

[10] W. F. Dearborn, *The Psychology of Reading*, Archives of Philosophy, Psychology and Scientific Methods, No. IV (1906), p. 131.

[11] W. C. Ruediger, *The Field of Distinct Vision*, Archives of Psychology, Vol. I, No. 5 (1907).

[12] H. R. Crosland, "Superior Elementary-School Readers Contrasted with Inferior Readers in Letter-Position, 'Range of Attention' Scores," *Journal of Educational Research*, XXXII (1939), 410–427.

Crosland, working with subjects at the elementary school level, used the technique of the tachistoscopic presentation of series of unrelated letters. He found, in a comparison of 31 superior with 34 inferior readers, that inferior readers excel superior readers in the accuracy of perception in the right-most positions of the visual field, whereas superior readers excel inferior readers in the accuracy of perception in the left-most positions.

The greater accuracy of inferior readers in their perception in the right field of vision, noted by Crosland, seems especially significant in view of the nature of the role which peripheral stimuli have been thought to play in the reading process. If a more adequate knowledge of the factors that influence the reader's use of peripheral stimuli is to be obtained, there is a manifest need for more data concerning how accuracy of perception in both the right and the left peripheral fields of vision is related to reading ability and to reading habits. It is the purpose of the present study to supply some of these data.

Procedure

The subjects for the experiment were sixty college freshmen, selected to simulate a normal or representative distribution with regard to intelligence. The data for each subject consist of scores obtained from: (1) seven tests of peripheral perception, (2) eye-movement records of the silent reading of both easy and difficult material, and (3) standardized tests for intelligence, spelling, vocabulary, and reading comprehension.

The tests of peripheral perception were administered individually, and consisted of recognizing in peripheral vision different types of materials when exposed tachistoscopically under constant conditions of duration and light intensity. The apparatus used for presenting these materials was an adaptation of the Whipple Disc Tachistoscope. The alterations made in this instrument accomplished several purposes. They made possible a stationary point of fixation in the pre-exposure which remained in view during the exposure, and they also kept the pre-exposure, exposure, and post-exposure of the same light intensity.

Exposure time for all tests was 93.75 sigma. The lighting was designed to make intensity optimal and measured 75 Weston.[13]

[13] Measured by the Weston Master Universal Exposure Meter, Model 715.

During each exposure the experimenter, by means of an adjustable mirror, watched the subject's eyes in order to be sure that they had not moved from the point of fixation. The distance of twelve inches between the subject's eyes and the point of fixation was kept constant by means of an adjustable head rest.

In constructing the tests for peripheral perception, the investigator chose materials presenting the least difficulty in reporting and permitting objective scoring. Also materials were selected which seemed not to introduce memory as an important variable. Seven types of test materials were selected as best suited for the study.

In the first three tests of peripheral perception, the items were dots, capital letters, and numerals, respectively. These test materials were presented simultaneously in both peripheral fields. In Test 1 from two to six dots were exposed simultaneously on the left and right sides of the point of fixation. The number of dots on either side never exceeded five; the total on both sides never exceeded six. In Test 2 two different capital letters were exposed simultaneously, one in each peripheral field. As in Test 1, the same items were presented in both the right and the left peripheral fields, but in different sequence. Test 3 was like Test 2, except that numerals were presented instead of capital letters.

In Tests 4, 5, 6, and 7, pairs of capital letters, simple words, pairs of words, and simple sentences of three words, respectively, were presented on either side of the point of fixation. In each of these tests half of the items were presented in the left peripheral field and half in the right, the side being determined by chance. In Test 4 two capital letters were exposed in the same peripheral field, either to the right or to the left of the fixation point.[14] Test 5 was composed of simple three-letter words, exposed one at a time, either to the right or to the left of the fixation point. Test 6 was composed of pairs of words; in many instances the pairs were phrases; in all instances the two words were related in such a manner as would permit their appearing in sequence in a sentence. All

[14] Test 4 was designed as a measure of peripheral perception which would permit variations in peripheral "masking" to enter as a variable. Because of the effect of masking in peripheral vision, two or more letters produce a mutual interference and can not be recognized as far from the point of fixation as can single letters. See R. S. Woodworth, *Experimental Psychology*, p. 720.

of the items in Test 7 were simple sentences of three words each.

The distance of the items from the point of fixation was increased at intervals in the first five tests. For all items in Tests 6 and 7, the distance from the fixation point to the letter in the item nearest it was the same. Each of the tests contained sixty test items; all test items were made with a typewriter, and each item was typed on a separate card. The sequence of the test items for each test, and the order in which the tests were taken were the same for all subjects. Subjects' scores on the tests were based on credit for items, or parts of items, perceived correctly with no credit for items perceived incorrectly or in incorrect sequence.[15] The reliabilities of the tests, as determined by the even-odd method, were as follows: Test 1, .813; Test 2, .916; Test 3, .898; Test 4, .898; Test 5, .890; Test 6, .924; and Test 7, .954.

Results

The results of the study concern the manner in which each of twenty-eight measures of peripheral perception is related to twenty-one measures of factors involved in or associated with reading. The twenty-eight measures of peripheral perception were of four kinds determined for each of the seven tests. These were: (1) total scores—these scores were based on the accuracy of perception for the tests as a whole and included items appearing in both peripheral fields; (2) scores for the right peripheral field—these scores represented the sum of points credited for successes on those items, or parts of items, appearing to the right of the fixation point; (3) scores for the left peripheral field—these were the sum of points credited for successes on those items, or parts of items, appearing to the left of the fixation point; and (4) right-field-dominance—these measures represented the score on the right expressed as a percentage of the total score.

The twenty-one measures of factors involved in, or associated with, reading were as follows: on the 1939 edition of the American Council on Education Psychological Examination for College Freshmen, there were three measures—the total, quantitative, and linguistic scores. Scores for spelling and vocabulary were obtained from Form PM of the American Council's English Test.[16] Measures involving speed, level of comprehension, vocabulary, and total scores were obtained from Form Q, Test C 1, of the Council's Test for Reading Comprehension. From the eye-movement records of the silent reading of both the easy and difficult material,[17] considered separately, scores were computed for: (1) reading time, (2) number of fixations, (3) average duration of fixations, (4) regression time, (5) number of regressions, and (6) average duration of regressions.

The relationships between these measures of peripheral perception and factors in reading were treated in three ways: (1) by a comparison of the upper[18] and lower quarter groups on each measure of peripheral perception with respect to their average[19] performance on each of the measures of factors in reading, (2) by a comparison of the upper and lower quarter groups on each measure of factors in reading with respect to their average performance on each of the measures of peripheral perception, and (3) in terms of correlations, for the experimental group as a whole, between the measures of peripheral perception and factors in reading.

For the upper and lower quarter groups on each of the measures of peripheral perception, the differences in average scores[20] made on each of the measures of factors in reading are given in Table I. Listed from top to bottom in Table I, the rows designate the measures of peripheral perception on which the sortings were made. For each of the seven tests of peripheral perception, the four types of measures are identified as total, right, left, and right-field-dominance respectively. Listed from left to right in Table I, columns 1, 2, and 3 show the differences in averages obtained for total scores, quantitative scores,

[15] In Test 1, success was based on the accuracy with which subjects recognized the correct number of dots; in all other tests subjects were required to discriminate the characters shown.

[16] All of the standardized tests used in the study are published by the American Council on Education.

[17] The selections representing the easy and difficult material were paragraphs 4 and 12 respectively of W. S. Gray's *Standardized Oral Reading Paragraphs*. These selections have been similarly used by Judd and Buswell in collecting data concerning the manner in which individuals adjust to reading material as its difficulty is increased. Cf. C. H. Judd and G. T. Buswell, *Silent Reading:* A Study of the Various Types. Supplementary Educational Monographs, No. 23. Chicago: University of Chicago Press, 1922.

[18] The term "upper quarter group" is used in all instances to mean the group receiving the largest scores on a particular measure, irrespective of whether the large score is one associated with reading ability or with reading disability.

[19] The term "average" is used to refer to arithmetic mean.

[20] Where the average of the upper quarter group was greater than that of the lower quarter group, the difference is recorded as positive; where the average of the upper quarter group was less than that of the lower quarter group, the difference is recorded as negative.

TABLE I

DIFFERENCES IN AVERAGES FOR FACTORS IN READING BETWEEN SUBJECTS WHO FORM THE UPPER AND LOWER QUARTER GROUPS ON MEASURES OF PERIPHERAL PERCEPTION

			Psychological Exam.			English Test		Reading Test			
			Total 1	Quan. 2	Ling. 3	Spelling 4	Vocab. 5	Total 6	Vocab. 7	Speed 8	Level 9
Test 1—Dots	Total	1	3.63	4.66	−1.02	−2.67	−1.54	−2.02	−2.44	1.16	−1.51
	Right	2	.64	1.65	−1.02	−5.37	−3.35	−.99	−.77	−2.33	−.30
	Left	3	−5.12	−1.16	−3.95	−.29	−1.37	−1.92	−.83	1.41	−.94
	Right Dom.	4	−.66	.70	−1.37	−4.00	−2.80	−.09	.75	−4.69	.85
Test 2—Capitals	Total	5	11.64	4.05	7.58	−1.28	−.51	2.11	.54	2.13	.48
	Right	6	5.23	.62	4.60	−2.55	−2.48	−1.12	−.23	−3.33	−3.10
	Left	7	8.33	2.29	6.04	4.03	4.08	4.14	2.11	5.27	4.46
	Right Dom.	8	−11.54	−4.90	−6.73	−7.47	−4.40	−5.87	−2.73	−8.33 3.46	−6.06 3.05
Test 3—Numerals	Total	9	1.57	1.37	.20	−1.43	−6.14	−1.02	−3.72	−.30	.71
	Right	10	−6.07	−.81	−5.25	−6.07	−9.31 3.71	−3.75	−2.88	−4.44	−3.62
	Left	11	20.89 7.10	8.50 3.13	12.39 4.71	11.45 4.05	3.90	4.34	−.76	6.93 3.36	4.85
	Right Dom.	12	−21.36 6.94	−7.47 3.06	−13.88 4.50	−7.70 3.74	−10.35 3.64	−9.23 3.03	−6.29	−10.12 3.08	−9.06 2.81
Test 4—Pairs of Capitals	Total	13	6.75	3.31	3.44	−.56	−3.20	−3.00	−5.25	.88	−1.69
	Right	14	−6.97	−2.33	−4.65	−3.61	−6.93	−6.86 3.09	−7.91 3.59	−6.43	−5.16
	Left	15	16.01 6.72	5.24	10.77 4.38	5.08	4.05	2.42	−1.02	4.88	3.33
	Right Dom.	16	−23.93 6.56	−8.10 2.91	−15.83 4.31	−10.30 3.63	−10.75 3.75	−12.24 2.95	−9.35 3.65	−13.14 3.10	−9.02 2.88
Test 5—Words	Total	17	16.52 7.31	6.16	10.36 4.77	6.57	2.39	1.09	.11	3.69	2.75
	Right	18	9.36	2.86	6.49	4.06	−.31	−.23	−1.09	2.64	1.00
	Left	19	15.46 7.13	6.55 3.14	8.90	6.86	3.29	1.52	−1.20	2.82	3.19
	Right Dom.	20	−10.97	−4.27	−6.70	−4.47	−4.51	−4.70	−1.17	3.94	5.27
Test 6—Pairs of Words	Total	21	18.47 7.55	6.13	12.33 4.93	8.60 4.05	2.06	1.47	1.00	2.86	.67
	Right	22	2.54	−.17	2.70	3.43	−2.10	3.37	−2.61	−1.31	−1.83
	Left	23	13.04	3.55	9.49	8.78 4.05	2.27	1.12	−3.08	2.74	−.38
	Right Dom.	24	−8.48	−1.96	6.53	4.14	−.62	−1.34	−2.39	−2.39	−2.33
Test 7—Sentences	Total	25	14.67 7.70	4.47	10.20 5.03	10.13 3.93	2.60	−.47	−.67	3.07	.80
	Right	26	3.06	−.63	3.70	4.57	1.75	−2.13	−1.28	−.15	−1.38
	Left	27	18.30 7.45	8.14 3.19	10.16 4.86	7.08	1.44	2.03	.41	3.05	1.60
	Right Dom.	28	−9.74	−4.56	−5.17	−.59	−3.15	−3.72	−1.45	−2.38	−2.42
	S. E. of Diff.		8.44	3.71	5.52	4.42	4.48	3.71	4.24	3.92	3.43

and linguistic scores, respectively, on the Psychological Examination. Columns 4 and 5 indicate the differences in average scores for spelling and vocabulary respectively. Columns 6, 7, 8, and 9 show the differences in averages for total scores, vocabulary scores, speed of comprehension scores, and level of comprehension scores, respectively, on the Reading Test. Columns 10, 11, 12, 13, 14, and 15 are all measures obtained from records[21] of eye-movements made by subjects when reading the easy material. These columns show respectively the total reading time, number of fixations, average length of fixations, total regression time, number of regressions, and the average length of regressions. Columns 16, 17, 18, 19, 20, and 21 show respectively these same measures for the difficult material.

[21] In all computations the first and last lines of each paragraph were omitted.

TABLE I—Continued

DIFFERENCES IN AVERAGES FOR FACTORS IN READING BETWEEN SUBJECTS WHO FORM THE UPPER AND LOWER QUARTER GROUPS ON MEASURES OF PERIPHERAL PERCEPTION

			Eye-Movement Measures—Easy Paragraph						Eye-Movement Measures—Difficult Paragraph					
			Read. Time	No. of Fixations	Avg. Length Fixations	Regr. Time	No. of Regres.	Avg. Length Regres.	Read. Time	No. of Fixations	Avg. Length Fixations	Regres. Time	No. of Regres.	Avg. Length Regres.
			10	11	12	13	14	15	16	17	18	19	20	21
Test 1—Dots	Total	1	10.71	.35	.28	−2.89	−.95	−.37	−15.15	−4.44	.21	−5.70	−1.66	.06
	Right	2	18.34	2.92	.19	−1.97	−.51	−.09	1.97	−.32	.10	.14	−.29	.75
	Left	3	2.28	−.78	.19	−1.11	−.34	.18	−16.69	−5.30	.21	−3.47	−1.33	.28 −.42
	Right Dom.	4	12.08	3.70	−.08	1.10	.26	−.22	5.98	2.24 3.13	−.19	1.07	.58	−.21
Test 2—Capitals	Total	5	4.42	2.08	−.21	5.44	1.02	.31	11.21	3.37	−.30	3.67	1.11	−.19
	Right	6	10.23	1.74	.02	−1.37	−.29	−.09	21.90	2.87	.07	−1.21	−.61	.20
	Left	7	−2.49	1.67	−.27	9.19	2.02	.32	−11.50	.35	−.41	4.06	1.16	.14
	Right Dom.	8	14.14	1.40	.22	4.29 −8.45	.95 −1.86	−.18	26.10	.03	.22 .23	−3.58	−1.05	.02
Test 3—Numerals	Total	9	6.05	1.17	.012	2.23	.30	.20	−10.57	.61	−.42	−3.30	−.82	.16
	Right	10	20.56	4.00	.05	−.31	−.25	.26	13.18	3.19	.21 −.16	.75	−.07	.15
	Left	11	−28.83	−2.85	−.45	.43	.26	−.31	−36.95	−3.57	−.55	−4.82	−.85	.15
	Right Dom.	12	13.40 32.18 13.09	6.24 2.42	.21 .09	4.23	.64	.45	15.01 21.00	3.94	.20 .005	3.11	.65	.03
Test 4—Pairs of Capitals	Total	13	−5.50	.81	−.27	4.07	1.00	−.08	−11.69	−1.00	−.27	2.69	.87	−.09
	Right	14	25.48 13.50	5.67 2.23	−.04	7.66	1.45	.35	15.38	3.24	−.08	4.75	1.29	−.15
	Left	15	−15.92	−1.55	−.25	2.58	.60	−.02	−31.26 14.90	−4.67 2.38	−.23	−2.98	−.16	−.27
	Right Dom.	16	40.04 12.21	6.23 2.17	.37	3.05	.76	−.11	40.00 14.30	4.85 2.25	.42 .24	2.40	.24	.23
Test 5—Words	Total	17	−7.93	−.69	−.13	3.65	.83	.12	−24.05	−3.44	−.25	−1.29	−.20	−.07
	Right	18	−2.28	.52	−.13	1.86	.37	.18	−13.84	−1.23	−.21	1.10	.06	−.02
	Left	19	−18.30	−3.05	−.08	1.05	.16	−.05	−29.90	−5.17 2.36	−.04	−5.02	−.96	−.11
	Right Dom.	20	14.08	3.69	−.08	−.97	.09	−.17	4.90	2.67	.02	2.04	.54	−.32
Test 6—Pairs of Words	Total	21	−22.93 14.15	−3.33	−.23	−.26	−.20	.12	−32.27 16.29	−2.46	−.56 .22	−.53	.20	−.18
	Right	22	−14.18	−2.59	−.08	−2.78	−.81	.24	−13.78	−.75	−.26	−.24	−.23	.07
	Left	23	−13.12	−2.74	−.21	1.74	.27	.02	−17.42	−1.63	−.30	.40	.52	−.35
	Right Dom.	24	12.70	1.37	.15	−3.43	−.79	.10	17.69	2.76	.14	.35	−.29	.20
Test 7—Sentences	Total	25	−18.34	−3.53	−.07	00	−.07	−.03	−29.53	−3.40	−.31	−.47	−.20	−.03
	Right	26	−15.52	−2.85	−.06	−4.86	−1.24	.21	−29.73	−1.20	−.27	−.33	−.41	.30
	Left	27	−.22	−1.53	.14	4.62	1.04	−.16	−8.07	−3.30	.15	−.46	−.27	.12
	Right Dom.	28	−2.72	−1.19	.13	4.68	1.27	−.34	5.90	−.40	.12	−.35	.36	−.32
	S. E. of Diff.		15.10	2.66	.23	5.06	1.11	.27	17.48	2.66	.24	4.82	1.11	.31

The standard error of the difference for chance samplings of fifteeen was determined for the data recorded in each of the columns in Table I. These are shown at the bottom of columns 1 to 21 inclusive. In all cases where the differences shown in the table were greater than 1.73 times this standard error, these differences were selected as those for which the actual standard errors of the differences were computed. In all cases where a difference shown in Table I is greater than two times its actual standard error, the standard error of the difference is recorded directly under the difference in averages.

In Table I large scores on those measures obtained from standardized tests (columns 1

to 9 inclusive) indicate either reading ability or success in factors associated with reading ability. On the other hand, high scores on those measures obtained from eye-movement records (columns 10 to 21 inclusive) are in no case associated with reading ability and in some instances indicate reading disability. Positive differences on columns 10 and 16, for example, correspond to negative differences on column 8, for with respect to column 8 the larger scores on the reading test were obtained by the more rapid readers, but with columns 10 and 16 the larger scores were obtained by the slower readers.

It can be noted from Table I that, except for Test 1, there is a very definite tendency in all tests for subjects who excel in the right peripheral field to be inferior in reading ability and in those factors associated with reading ability, and for subjects who excel in the left peripheral field to be consistently superior in all of these factors. In fact, in all of the comparisons shown in Table I there is not (excluding Test 1 and the measures for regressions[22]) a single instance in which success in the left peripheral field is not more positively related to reading ability and the factors associated with it than is success in the right peripheral field. In most instances success in the right peripheral field is actually negatively related to factors associated with reading ability. In terms of the measure of right-field-dominance, in every instance (excluding Test 1 and the measures for regressions) those subjects forming the upper quarter group on this measure were inferior to the lower quarter group in reading ability and in factors associated with reading ability.

The largest differences for these various factors in reading were for measures on Tests 3 and 4. On these tests particularly may be seen the opposite manner in which the accuracy of perception in the right and left fields of vision is related to factors associated with reading ability. The differences listed for Tests 3 and 4 in rows 10 and 14, for example, show that in terms of success on the right the upper quarter groups had lower average scores on the psychological examination than the lower quarter groups; the differences listed in rows 11 and 15 show that in terms of success on the left the upper quarter groups had higher average scores on the psychological examination than the lower quarter groups; the differences listed in rows 12 and 16 show that in terms of right-field-dominance the upper quarter groups had lower average scores on the psychological examination than the lower quarter groups.

A comparison of the two upper quarter groups (the upper quarter group for success on the left and the upper quarter group for scores on right-field-dominance) shows still greater differences on the various measures of factors in reading than those indicated in the table. For these two upper quarter groups on Test 3, for example, the difference in average scores made on the psychological examination was 25.03 (S. E. of difference 6.32); the difference in average scores made on spelling was 13.03 (S. E. of difference 3.62); the difference in average scores made on speed of reading was 42.04 (S. E. of difference 12.03). These relationships will be presented and discussed later in terms of correlations.

The consistency of certain aspects of the data given in Table I seems to give to some of the findings concerning the relationships between variations in peripheral perception and factors in reading a reliability far greater than can be indicated in terms of the standard error of the difference for any particular comparison. A number of differences, however, are large enough to be statistically significant. These are easily identified in Table I, because of the fact that in all cases where a difference is greater than two times its standard error, the standard error of the difference is shown directly under the difference.

The differences shown in Table II were determined in the same manner, and for a study of the same relationships as those shown in Table I, except that in Table II are given for the upper and lower quarter groups, on each of the measures of factors in reading, the differences in average scores made on each of the measures of peripheral perception. From left to right the columns show the differences in average scores on the measures of peripheral perception. From top to bottom the rows show the factors in reading on which the upper and lower quarter groups were selected. At the bottom of columns 1 to 28 inclusive is shown, for the data recorded in each of the columns, the standard error of the difference based on chance samplings of fifteen. In Table II, as in Table I, in all instances where the differences shown

[22] Whether regressions within limits are indicative of or associated with reading disability, and should therefore be mentioned as an exception, is open to question.

in the table were greater than 1.73 times this standard error, these differences were selected as those for which the actual standard errors were computed. Where a difference shown in Table II is larger than two times its actual standard error, the standard error is shown directly under the difference.

The same tendencies noted in Table I can be seen also in Table II, namely, a tendency for superiority in the accuracy of perception in the left peripheral field to be associated with factors indicative of reading ability, and a tendency for superiority in the accuracy of perception in the right peripheral field and high scores on right-field-dominance to be associated with factors indicative of reading disability. For example, in every instance, except for Test 1 and the measures of regressions, the upper quarter group on any factor associated with reading ability had a lower average on right-field-dominance. Also, in nearly every instance the upper quarter group on any factor associated with reading ability had a higher average for success in the left peripheral field. For Tests 2, 3, and 4, in nearly every instance, the upper quarter group on any factor associated with reading ability was inferior to the lower quarter group in the accuracy of their perception in the right peripheral field. In Tests 5, 6, and 7, the upper quarter groups in some instances were superior in the right as well as in the left peripheral field, but the superiority in the right was in no instance as great as the superiority in the left. The relationships found between factors in reading and all of the measures of peripheral perception, except those for total scores, exhibit a consistency which on the basis of chance would occur very rarely.

TABLE II

DIFFERENCES IN AVERAGES FOR MEASURES OF PERIPHERAL PERCEPTION BETWEEN SUBJECTS WHO FORM THE UPPER AND LOWER QUARTER GROUPS ON FACTORS IN READING

			Test 1—Dots				Test 2—Capitals				Test 3—Numerals				Test 4—Pairs of Capitals			
			Total 1	Right 2	Left 3	Right Dom. 4	Total 5	Right 6	Left 7	Right Dom. 8	Total 9	Right 10	Left 11	Right Dom. 12	Total 13	Right 14	Left 15	Right Dom. 16
Psych. Exam.	Total	1	3.50	1.50	2.00	—.25	2.87	—2.00	4.87	—5.12	4.19	—2.62	6.81	—6.06	3.63	—2.68	6.31	—8.44
													2.14	1.82			2.07	2.35
	Quan.	2	2.63	1.13	1.50	—.29	3.58	—.85	4.43	—4.04	3.86	—1.72	5.58	—5.08	4.08	—1.37	5.45	—6.99
													2.17	1.82				2.37
	Ling.	3	1.66	.94	.74	.27	2.00	—1.67	3.66	—3.73	1.06	—5.14	4.07	—6.47	1.20	—3.86	5.06	—7.54
												2.17		1.80				2.37
Eng. Test	Spelling	4	—3.02	1.40	—1.61	.34	1.36	—2.24	.89	—2.82	.25	—1.58	1.84	2.08	—1.15	—3.75	2.60	—6.41
																		2.48
	Vocabulary	5	—1.61	—1.81	.28	—1.74	—1.13	—2.52	1.38	—3.04	—2.30	—5.31	3.01	—5.69	—5.01	—5.89	.87	—7.00
												2.01		1.80				2.53
Reading Test	Total	6	—3.00	—1.60	—1.40	—.13	3.60	—2.60	6.20	—6.74	—1.67	—5.86	4.20	—7.07	—.40	—4.20	3.80	—8.07
									2.61			2.12		1.81		1.84		2.41
	Vocabulary	7	—1.22	.72	—1.95	1.70	—.61	—2.77	2.17	—3.79	—2.24	—2.98	.75	—3.04	—.95	—3.45	2.51	—5.82
																		2.60
	Speed	8	—2.57	—2.14	—.44	—1.04	2.33	—2.11	4.45	—4.47	.17	—4.66	4.83	—6.52	—.60	—5.30	4.72	—9.12
												2.10	2.17	1.73		1.90	2.26	2.39
	Level	9	—1.08	.02	—1.09	.62	4.57	—1.17	5.74	—5.11	.72	—2.52	3.25	—4.46	2.50	—2.51	5.01	—7.41
									2.58					1.42			2.37	2.54
Eye-movement measures Easy paragraph	Reading Time	10	.20	1.34	—1.13	1.93	—.87	.73	—1.60	1.60	.53	4.80	—4.26	5.34	—1.28	3.40	—4.67	8.20
												2.22		1.91				2.45
	No. of Fixations	11	—2.90	.24	—3.13	2.61	2.23	.04	2.19	—1.81	3.81	3.82	00	1.87	3.52	4.95	—1.43	6.49
																1.84		2.45
	Avg. length of fixations	12	.69	.69	00	1.07	—5.50	.93	—6.44	5.44	—7.63	—.37	—7.25	4.31	—8.38	—1.31	—7.06	5.00
									2.76		3.47		2.33		3.49		2.33	
	Regression time	13	—3.54	—.46	—3.07	1.86	6.80	1.07	5.73	—3.34	1.93	2.00	—.07	.87	5.26	4.20	1.06	3.74
									2.47							2.00		
	No. of Regressions	14	—5.50	—1.67	—3.82	1.98	2.59	—2.15	4.74	—5.42	.65	—.30	.95	—1.26	2.40	2.53	—.12	3.09
	Avg. Length of Regression	15	1.04	.59	1.63	1.45	2.98	.03	2.95	—2.23	—1.72	.84	—2.56	1.58	—1.69	—.86	—.83	—.89
Eye-movement measures Difficult paragraph	Reading Time	16	—4.26	—1.47	—2.80	.80	—.87	1.07	—1.94	1.53	—5.66	.53	—6.20	3.60	—4.20	2.00	—6.20	6.93
													2.18				2.21	2.65
	No. of Fixations	17	—5.73	—2.47	—3.27	.34	—1.00	—.07	—.93	.66	.67	2.60	—1.94	2.46	—2.60	2.53	—5.13	7.67
																	2.33	2.56
	Avg. Length of Fixations	18	5.46	2.73	2.74	.20	—9.47	—6.13	—3.33	2.07	—7.07	—1.40	—5.66	2.67	—6.66	—.33	—6.33	4.26
											3.08		2.13				2.26	
	Regression time	19	—3.46	—1.03	—1.54	—.33	—1.93	—4.20	2.27	—5.53	—1.13	1.13	—2.27	1.27	—2.20	—.27	—1.93	—.07
	No. of Regressions	20	—3.65	—1.47	—2.18	.46	.88	—1.42	2.31	—2.78	—1.75	.89	—2.65	1.48	2.93	2.81	.11	2.65
	Avg. Length of Regression	21	3.33	1.81	1.52	.29	—1.85	1.17	—3.03	3.08	—1.31	1.22	—2.53	2.34	—5.63	—2.39	—4.26	—.33
	S. E. of Diff.		3.87	2.23	2.52	1.98	3.95	2.88	2.92	3.30	3.59	2.42	2.53	2.21	3.81	2.21	2.59	3.02

TABLE II—Continued

DIFFERENCES IN AVERAGES FOR MEASURES OF PERIPHERAL PERCEPTION BETWEEN SUBJECTS WHO FORM THE UPPER AND LOWER QUARTER GROUPS ON FACTORS IN READING

			Test 5—Words				Test 6—Pairs of Words				Test 7—Sentences			
			Total 17	Right 18	Left 19	Right Dom. 20	Total 21	Right 22	Left 23	Right Dom. 24	Total 25	Right 26	Left 27	Right Dom. 28
Psych. Exam.	Total	1	9.88	1.50	8.37	−6.25	10.88	1.25	9.63	−8.44	13.56	2.12	11.44	−6.31
			4.25		2.67	2.89	4.92		3.28	3.46	6.52		4.71	
	Quan.	2	7.92	1.67	6.25	−4.32	6.65	1.49	5.17	−3.73	8.12	−.08	8.21	−5.17
					2.80									
	Ling.	3	7.60	1.34	6.27	−4.34	9.00	1.27	7.73	−7.13	8.93	.80	8.14	−5.00
					2.80				3.50	3.42				
Eng. Test	Spelling	4	3.54	1.00	2.54	−1.67	8.05	3.29	4.75	−3.96	16.09	7.82	8.27	−1.56
											6.33			
	Vocabulary	5	−.68	−2.87	2.18	−3.08	−3.65	−4.62	.98	−4.60	−3.54	−6.04	2.50	−4.51
Reading Test	Total	6	4.40	00	4.40	−4.87	2.40	−2.34	4.73	−6.40	−.40	−5.14	4.74	−5.93
	Vocabulary	7	−.42	−.77	.36	−.74	−.85	−.12	.28	−1.54	−2.28	−3.94	1.66	−2.69
	Speed	8	4.22	.31	3.92	−2.62	4.15	−.26	4.42	−4.58	2.07	−2.33	4.40	−4.06
	Level	9	7.93	1.82	6.13	−5.37	.93	−2.52	3.46	−4.34	1.17	−3.10	4.28	−4.45
					2.80	2.82								
Eye-movement measures Easy paragraph	Reading Time	10	−4.00	00	−4.00	3.06	−3.66	−.66	−3.00	.46	−3.07	−1.46	−1.60	2.54
	No. of Fixations	11	−1.08	.82	−1.90	1.17	−4.06	−3.64	−.41	−.88	−5.74	−4.48	−1.26	−.92
	Avg. length of fixations	12	−2.88	−1.00	−1.87	1.37	−6.50	.25	−6.75	6.69	−6.00	−.75	−5.25	1.75
									3.69					
	Regression time	13	2.93	1.73	1.20	−1.46	2.00	−.74	2.73	−2.40	−2.60	−3.80	1.20	−3.16
	No. of Regressions	14	.89	−1.85	2.74	−5.06	−.42	−4.56	4.15	−6.15	−1.25	−6.69	5.45	−6.56
	Avg. Length of Regression	15	−5.65	−2.45	−3.20	.64	−7.51	−1.89	−5.63	5.14	−12.39	−3.66	−8.72	2.70
Eye-movement measures Difficult paragraph	Reading Time	16	−6.80	−1.87	−4.93	1.93	−3.86	−.27	−3.60	3.67	−9.06	−2.87	−6.20	1.33
	No. of Fixations	17	−8.13	−.53	−7.60	6.46	−1.87	2.40	−4.26	5.93	−5.06	3.27	−8.33	6.13
					2.81	2.94								
	Avg. Length of Fixations	18	−5.47	−1.87	−3.60	1.07	−8.20	−1.94	−6.27	4.86	−7.74	−4.93	−2.80	−.93
	Regression time	19	−7.20	−2.94	−4.26	1.60	−.40	−1.20	.80	−1.00	−5.00	−2.20	−2.80	−1.00
	No. of Regressions	20	−2.56	−.12	−2.45	1.51	1.96	−.65	2.61	−2.51	−4.39	−2.98	−1.40	−1.09
	Avg. Length of Regression	21	−3.56	−1.23	−2.34	.79	−3.50	1.22	−4.72	6.17	−3.37	3.05	−6.41	4.64
	S. E. of Diff.		4.93	2.91	3.17	3.14	5.57	3.63	3.82	3.82	7.59	5.71	5.02	3.83

In addition to the characteristic of consistency, the reliability of many of the findings shown in Table II, as in Table I, is indicated by differences in averages larger than two times their standard error. It can be noted, for example, that for total scores on the psychological examination (row 1) the differences in averages on twelve measures of peripheral perception were larger than two times their standard error. In Table II, as in Table I, it is the measures on Tests 3 and 4 which show the largest number of statistically significant differences. For almost all of the factors associated with reading ability, reliable differences can be seen in the case of right-field-dominance on these tests.

Relationships between the measures of peripheral perception and factors in reading which were treated in terms of zero-order correlations are shown in Tables III and IV.[28] Not all possible correlations were computed, because the differences in averages, shown in Tables I and II, indicated that those correlations omitted would have been found too low to be reliable.

In Table III the correlations between some of the factors in reading and total scores, scores on the right, and scores on the left for each of the tests respectively are given. From top to bottom, the factors listed appear in the following order: (1) total scores on the psycholgical examination, (2) scores for spelling on the English test, (3) total scores on the reading test, (4) scores for speed on the reading test, (5) scores for level on the reading test, (6) scores for total reading time on

[28] The correlations shown in these tables that are less than .28 have a probable error of ±.08; those larger than .28 have a probable error of ±.07.

TABLE III

RELATIONSHIPS IN TERMS OF CORRELATION OF FACTORS IN READING TO MEASURES OF PERIPHERAL PERCEPTION

		Test 1—Dots			Test 2—Caps.			Test 3—Numerals			Test 4—Pairs		
		Total 1	Right 2	Left 3	Total 4	Right 5	Left 6	Total 7	Right 8	Left 9	Total 10	Right 11	Left 12
Psy. Exam. Total	1	.040	.051	−.000	.211	−.065	.225	.102	−.154	.291	.151	−.173	.367
English Test Spelling	2	−.099	−.142	−.058	.035	−.130	.181	.021	−.212	.233	.083	−.147	.252
Reading Test Total	3	−.052	−.038	−.091	.097	−.066	.024	−.027	−.233	.183	−.065	−.306	.173
Reading Test Speed	4	−.007	−.079	.019	.108	−.098	.251	.024	−.244	.267	.008	−.259	.240
Reading Test Level	5	−.058	−.130	−.007	.083	−.101	.220	.011	−.200	.207	−.016	−.216	.167
Easy Parag. Read. Time	6	.051	.122	−.072	.037	.072	−.029	−.040	.162	−.211	−.042	.200	−.226
Easy Parag. No. Fixa.	7	−.011	.091	−.140	.113	.083	.078	.039	.175	−.111	.070	.305	−.129
Diff. Parag. Read. Time	8	−.087	−.040	−.145	.033	.127	−.148	−.150	.053	−.263	−.138	.092	−.273
Diff. Parag. No. Fixa.	9	−.208	−.088	−.291	.090	.152	−.020	−.021	.114	−.139	−.059	.139	−.179

TABLE III—Continued

RELATIONSHIPS IN TERMS OF CORRELATION OF FACTORS IN READING TO MEASURES OF PERIPHERAL PERCEPTION

		Test 5—Words			Test 6—Pairs of Words			Test 7—Sentences		
		Total 13	Right 14	Left 15	Total 16	Right 17	Left 18	Total 19	Right 20	Left 21
Psy. Exam. Total	1	.251	.118	.288	.201	.074	.226	.168	.027	.222
English Test Spelling	2	.203	.117	.214	.234	.116	.235	.304	.196	.236
Reading Test Total	3	.038	−.054	.116	−.056	−.131	.048	−.075	−.131	.035
Reading Test Speed	4	.137	.047	.177	.026	−.062	.122	.000	−.057	.065
Reading Test Level	5	.152	.030	.216	−.063	−.112	.019	−.047	−.088	.029
Easy Parag. Reading Time	6	−.083	.000	−.126	−.120	−.055	−.119	−.126	−.107	−.068
Easy Parag. No. Fixa.	7	−.050	.036	−.107	−.113	−.107	−.058	−.137	−.136	−.049
Diff. Parag. Read. Time	8	−.157	−.056	−.051	−.130	−.037	−.151	−.180	−.109	−.148
Diff. Parag. No. Fixa.	9	−.164	−.044	−.211	−.062	−.002	−.089	−.147	−.050	−.165

the easy paragraph, (7) scores for total number of fixations on the easy paragraph, (8) scores for total reading time on the difficult paragraph, and (9) scores for total number of fixations on the difficult paragraph.

Most of the correlations shown in Table III are too small to be statistically reliable when considered separately. The correlations are all based on sixty cases, and all of these correlations less than .25 are less than three times their probable errors. The consistency that may be noted in the correlations, however, is a characteristic which should not be overlooked in evaluating their statistical significance. For example, it can be seen in the case of Tests 2, 3, and 4 that the relationships between success on the right and factors associated with reading ability are in every instance negative.[24] Also, in Tests 2, 3, 4, 5, 6, and 7, with only two exceptions (row 7, column 6, and row 8, column 15), the relationships between success on the left and factors associated with reading ability are all positive.

Because of the tendency, particularly in Tests 3 and 4, for factors associated with reading ability to be negatively correlated with success on the right and positively correlated with success on the left, there are a few comparatively large as well as statistically reliable differences between these correlations. For example, the difference between these correlations for Test 4 (columns 11 and 12) and total scores on the psychological examination (row 1) is .540 ± .10; for Test 3 (columns 8 and 9) and speed of reading (row 4) it is .511 ± .10; for Test 4 (columns 11 and 12) and speed of reading (row 4) it is .499 ± .11; for Test 4 (columns 11 and 12) and reading time on the easy paragraph (row 6) it is .426 ± .12; for Test 4 (columns 11 and 12) and number of fixations on the easy paragraph (row 7) it is .434 ± .12.

The correlations between right-field-dominance on each of the tests except Test 1 and factors in reading are given in Table IV. The tests are listed in order from top to bottom. The factors in reading are listed from left to right. Many of the correlations shown in this table are statistically reliable. The consistency with which right-field-dominance is associated with factors indicative of reading disability is shown by the fact that, out of the sixty correlations in Table IV, there are only two in which the relationships between right-field-dominance and factors associated with reading ability are not negative, and in these two the correlations

[24] For rows 6, 7, 8, and 9 positive correlations, like positive differences in averages, represent negative relationships, since high scores indicate reading disability.

TABLE IV

RELATIONSHIPS IN TERMS OF CORRELATIONS OF RIGHT FIELD-DOMINANCE TO FACTORS IN READING

	Psy. Exam Total	Psy. Exam. Ling.	English Test Spell.	Reading Test Total	Reading Test Speed	Reading Test Level	Easy Para. Reading Time	Easy Para. No. Fixa.	Dif. Para. Reading Time	Dif. Para. No. Fixa.
	1	2	3	4	5	6	7	8	9	10
Right-Dom. Test 2	—.097	—.089	—.210	—.179	—.222	—.206	.055	.008	.078	.117
Right-Dom. Test 3	—.321	—.335	—.288	—.331	—.386	—.324	.248	.184	.190	.165
Right-Dom. Test 4	—.395	—.388	—.330	—.367	—.378	—.292	.345	.343	.233	.276
Right-Dom. Test 5	—.156	—.142	—.129	—.125	—.093	—.194	.084	.068	.017	.127
Right-Dom. Test 6	—.179	—.199	—.197	—.133	—.181	—.068	.120	.031	.110	.114
Right-Dom. Test 7	—.169	—.129	—.058	—.131	—.100	—.087	—.018	—.028	.001	.092

are so small that no relationships are indicated; they are —.018 and —.028 between right-field-dominance on Test 7 and reading time (column 7) and number of fixations (column 8) for the easy paragraph.

Mention has been made of the fact that for Tests 3 and 4 a comparison of the two upper quarter groups (the upper quarter group for success on the left and the upper quarter group for scores on right-field-dominance) shows still greater differences on the various measures of factors in reading than a comparison of the upper and lower quarter groups (either for success on the left or for scores on right-field-dominance). From a comparison of the correlations given in Tables III and IV, it can be seen that for Tests 3 and 4 the largest differences in correlations are between those for success on the left and for scores on right-field-dominance. For example, the correlation between total scores on the psychological examination and success on the left in Test 4 was .367; the correlation between total scores on the psychological examination and scores on right-field-dominance on Test 4 was —.395, making a difference of .762 ± .09, between the two correlations.

A summary of results shown in Tables I, II, III, and IV can be most conveniently grouped under the four types of measures of peripheral perception, in the following order: (1) right-field-dominance, (2) success on the left, (3) success on the right, (4) success on the totals.

(1) *Right-field-dominance.* — The highest and most consistent relationships between the measures of peripheral perception and the factors in reading were obtained for scores on right-field-dominance. In nearly every instance, in terms both of differences in averages and of correlations, right-field-dominance was negatively related to all factors associated with reading ability. These negative relationships were highest for the scores on right-field-dominance based on Tests 3 and 4; for these tests the following statistically reliable relationships were found. Subjects with high scores on right-field-dominance tended to have lower scores on the psychological examination, lower scores on spelling and vocabulary on the English Test, and lower scores on all measures on the standardized reading test. Relationships between right-field-dominance and scores based on eye-movement records showed that subjects with high scores on right-field-dominance were significantly slower in reading both easy and difficult material and that they made more and longer fixations.

(2) *Success on the left.*—Accuracy of perception in the left peripheral field was found to be related in a positive manner to nearly every factor involved in and associated with reading ability. In Tests 3 and 4 many of these relationships are large enough to be statistically reliable, and the consistency with which this finding can be noted in all tests seems to add appreciably to its statistical significance.

Subjects with high scores for success in recognizing items in the left peripheral field tended to have higher scores on the psychological examination, higher scores on spelling and vocabulary, and higher scores on all measures on the standardized reading test. Relationships between success on the left and scores based on eye-movement records showed that subjects with high scores for success on the left tended to be faster readers with fewer and shorter fixations.

(3) *Success on the right.*—In Tests 2, 3, and 4 success in recognizing items in the right peripheral field was found to be negatively related to all factors involved in and associated with reading ability. In Tests 5, 6, and 7 success on the right was in many instances slightly positively related to these factors, but in no case was the positive relationship as high as that for success on the left.

As was the case with the measures for right-field-dominance and success on the left, the largest number of statistically significant relationships for success on the right was found for Tests 3 and 4. Subjects with high scores for success on the right in these tests tended to make lower scores on the psychological examination, on spelling and vocabulary, and on all measures on the reading test. The relationships obtained for scores based on eye-movement records showed that subjects with high scores for success on the right in these tests tended to be slower readers with more and longer fixations.

(4) *Total scores.*—The smallest number of significant relationships, as well as the least consistent relationships, were found for total scores. In Tests 2, 3, and 4 the relationships for total scores were in most instances lower than those found in the case for both the right and the left sides, and in the remainder of the tests the relationships for scores based on one side tended to be larger than for total scores. This is to be expected in view of the opposite ways in which accuracy of perception in the right and left peripheral fields tends to be related to factors in reading.

Conclusions

The results of the present investigation have shown that, for the group studied, inferior readers excel superior readers in the accuracy of perception in the right peripheral field, but that superior readers excel inferior readers in the accuracy of perception in the left peripheral field. These relationships exist in the case of extremes, not only with respect to reading ability and related factors, but also with respect to the accuracy of perception in the right and left peripheral fields; and these same relationships exist for the experimental group as a whole.

A number of the findings indicate that the primary factor involved in producing these relationships is that of preference for the right or left peripheral fields and not that of variations in visual acuity or extent of clear vision.

The highest and most consistent relationships in the study were obtained between factors in reading and the measure designated as right-field-dominance. Scores on this measure represented a subject's accuracy in the right peripheral field in comparison with his accuracy in the left peripheral field; high scores indicated not superiority over other subjects, but rather a greater degree of preference for this field.

Little or no relationship was found between the accuracy of a subject's perception in his right field of vision and that of his left field of vision,[25] or between accuracy of perception in both peripheral fields and factors in reading. The relationships between accuracy of perception, as measured, and factors in reading seem to exist primarily when accuracy is indicative of preference.

The highest negative relationships between accuracy in the right field and factors in reading were found for those tests that were least like reading. For the group as a whole, these negative relationships were in Tests 3 and 4, where the test materials consisted of numerals and pairs of letters, whereas the lowest negative relationships were in Tests 6 and 7, where the test materials consisted of pairs of words and simple sentences. This finding may be attributed to self-correlation. Poor readers could show a greater superiority for items in the right field in Tests 3 and 4 than they could for items in the right field in Tests 6 and 7, because recognizing items in Tests 6 and 7 was more like reading. In fact, without some ability in reading, a subject could not have recognized any of the items in the right field in Tests 6 and 7.

The fact that it was in the left peripheral field where the superior reader's greater accuracy of perception was manifest, even in the case of pairs of words and simple sentences, might seem to indicate that recognizing pairs of words and sentences in the left peripheral field is more like reading than recognizing pairs of words and sentences in the right peripheral field. However, the item analyses of Tests 6 and 7 showed that, for the group as a whole, recognizing pairs of words and sentences in the right visual field was much easier than recognizing like pairs of words and sentences in the left visual field.[26]

[25] The correlations between success on the left and success on the right for each of the tests were as follows: Test 1, .274; Test 2, .057; Test 3, .061; Test 4, .244; Test 5, .283; Test 6, .122; and Test 7, .001.

[26] In Tests 1, 2, 3, and 4 items or parts of items presented in the left peripheral field were, for the group as a whole, easier to recognize. In Tests 5, 6, and 7 the recognition of items appearing in the right peripheral field was easier for the group as a whole. In all tests, however, there was considerable variability in the difficulty of particular test items. In all tests some items were missed three times as frequently as other items of the same type on the same side and at equal distance from the point of fixation. In the first five tests in which test items were presented at varying distances from the point of fixation, some items, even though twice as far from the point of fixation, were perceived correctly more frequently than other items of the same type on the same side.

That preference for the left peripheral field is associated with reading ability and related factors, and that preference for the right peripheral field is associated with reading disability and related factors, are conclusions which, except for two groups of findings, are confirmed by every aspect of the data in the present study. These two groups of findings are not incompatible with the conclusions; they merely fail to support them. They concern the lack of any substantial relationships found between: (1) measures based on Test 1 and factors in reading, and (2) measures based on all tests and regressive movements in reading.

The relationships between factors in reading and the measures for Test 1 were neither large nor consistent, and success on this test was only slightly related to success on any of the other tests.[27] If variations in success on Test 1 indicated differences in retinal sensitivity outside the fovea, then, so far as the group studied is concerned, this factor is not significantly related either to the accuracy of perception on the other tests or to factors in reading.

The relationships of regressive movements in reading to success in the left and right peripheral fields or to scores on right-field-dominance were too small to be statistically significant. This may have resulted from the fact that the number of regressions was too small to permit reliable measurement. On the other hand, it may be that regressive movements in reading, because they can result from such widely varied causes, are less related to either reading ability or reading disability than were the other measures.

In assuming that the relationships found in the present investigation may be attributed primarily to preference, the question arises as to why inferior readers, more than superior readers, prefer the right peripheral field (and therefore tend to be more accurate in recognizing items in it), while superior readers, more than inferior readers, prefer the left peripheral field (and therefore tend to be more accurate in recognizing items in it). This question takes on added significance, because of the importance in reading that has been attributed to vision in the right peripheral field. It has been assumed that in the reading process peripheral vision plays the role of forging ahead, thereby furnishing partial perceptions that contribute to efficient reading. Little emphasis has been given to peripheral perception as it might function in the left peripheral field in re-examinations or in filling in gaps that may have been passed over in direct vision.

That perceptions in the left peripheral field may function in the reading process rather significantly for some individuals seems to be indicated by the fact that a few of the subjects in the investigation were actually more accurate at reading sentences in the left peripheral field than they were at reading similar sentences in the right peripheral field. For the group as a whole, however, the very pronounced tendency for pairs of words and sentences to be easier when presented to the right of the point of fixation than when presented to the left indicates that normally it is the perception in the right peripheral field that predominates in the reading process.

The view that it is perception in the right peripheral field which functions more significantly in reading would seem to indicate that preference for the left peripheral field is not something that one acquires as the result of learning to read or as a result of a great amount of reading. If any cause and effect relationship can be assumed to exist between reading disability and preference for the right peripheral field, it would seem that it is the preference for the right peripheral field, or the condition of which this preference is symptomatic, which operates to produce the reading disability.

The work of Anderson and Crosland[28] shows that preference for the right or left fields of vision is related to "eyedness". These investigators, working with subjects at the college level and using the technique of the tachistoscopic presentation of series of unrelated letters, found that right-eyed individuals were superior in reporting letters in the left-most position of the visual field and left-eyed individuals were superior in reporting letters in the right-most positions. There

[27] Although the inter-correlations between success on Test 1 and the other tests were all low, inter-correlations of above .60 for all four types of measures of peripheral perception were found between some of the other tests in which were present such differences as: (1) the presentation of different types of test items, (2) the presentation of material in both peripheral fields and in separate peripheral fields, and (3) the presentation of test items at constant and at varying distances from the point of fixation. For example, in terms of scores for the right peripheral field, Test 4 correlated .632 with Test 2, .601 with Test 3, .711 with Test 5, and .627 with Test 6.

[28] I. H. Anderson and H. R. Crosland, "The Effect of Eye-Dominance on 'Range of Attention' Scores," *University of Oregon Publications*, Vol. VIII (1933), No. 4.

is therefore a rather striking consistency in the positive relationships between left-eyedness and reading disability, between left-eyedness and preference for the right peripheral field, and between preference for the right peripheral field and reading disability. However, if it is assumed that left-eyedness is associated with reading disability because of crossed cerebral dominance,[20] then it would seem that preference for the right peripheral field can not be assumed to be associated with reading disability for the same reason, since objects seen in the right peripheral field are imaged on the left half of each retina and are connected neurally with the left cerebral hemisphere.

It appears, therefore, that the inferior reader's greater accuracy of perception in the right peripheral field results from his greater preference for this field, but whether this preference is symptomatic of a condition, perhaps central in origin, which operates to produce reading disability, or whether it is a function of "set" acquired by the inferior reader as the result of his experiences with reading, is a problem that seems to require additional data before an entirely satisfactory interpretation can be formulated. Data concerning the relationships between accuracy of perception in peripheral vision and measures obtained from non-language tests would be of value in indicating possible causal relationships. Also, if the inferior reader's greater accuracy of perception in the right peripheral field results from his experiences with reading, then this hypothesis might be tested experimentally by obtaining data concerning: (1) the presence or absence of this relationship in the case of children whose reading habits have not been established, or (2) the presence or absence of a reversed relationship in the case of reading Arabic or Hebrew.

[20] Many writers have emphasized the point that it is where left-eyedness is accompanied by right-handedness that its association with reading disability is most evident.

EFFICIENCY IN LEARNING: AN EXPERIMENTAL STUDY OF LEARNING BY MEMORIZATION, BY A VERBAL RULE, AND BY NON-VERBAL CLUES TO UNDERSTANDING*

J. C. HEWSON
Superintendent of Schools, Castor School Division
Castor, Alberta, Canada

The Problem.—Recent literature on the psychology of learning has displayed an encouraging tendency to emphasize the points of agreement rather than the differences among learning theories. Insight was formerly taken to mean immediate solution of a problem without any tentative exploration, and many disputes have raged over its role in learning. There now seems to be general agreement that insight should be described as the perception of essential relationships in the learning situation, that it may occur in greater or lesser degree along with trial-and-error learning, and that it may be "foresight" or "hindsight" (1, 2, 3).

With this agreement on the crucial role of insight in learning, particularly in that kind of learning called problem solving, the way is cleared for constructive research in finding the best means of gaining insight in learning situations. This study attempts to test at three levels of maturity the relative efficiency of three methods of learning in aiding the learner to gain insight in a learning situation in which a complex mental problem must be solved. Specifically, the study attempts to answer experimentally the following questions:

1. Are memorizing, the use of a simple verbal principle or rule, and the use of non-verbal clues that may aid the learner to formulate the principles of the problem, efficient methods of learning to solve problems?

2. What is the relative efficiency of these methods at the junior high school level, the senior high school level, and the college level, as measured by the ability to make an immediate transfer to new but similar problems?

3. What is the relative efficiency of these methods at each level as measured by the retention of learned material?

4. What is the relative efficiency of these methods as measured by the ability to make a delayed transfer to new problems?

5. Is the mental age of the learner a more important factor in learning than the method of learning?

The Experimental Material.—The learning material used in the experiment consisted of nine mathematical puzzles, one of which is shown in Figure 1 with one of its four possible solutions. The problem is to change a

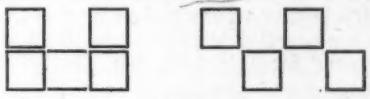

Figure 1
Mathematical Puzzle and Solution

figure of five equal squares made from sixteen lines to a figure of four equal squares of the same size by changing the position of three lines. Seven of the nine problems were figures of five squares, while the other two were figures of seven and nine squares to be changed to figures of five and seven squares respectively.

The Subjects.—The subjects were 689 students chosen in nearly equal numbers from a grade eight class in a junior high school, a grade eleven class in a large senior high school, and a college class. The school subjects at each level were divided into four groups equated in mean IQ and chronological age and of approximately equal variability. The college subjects were members of three sections of the same course in educational psychology and were also divided into four groups. Intelligence scores available for a large proportion of them indicated that the three groups used as learning groups were approximately equal in mental ability, while the control group was slightly lower than the learning groups.

The Experimental Procedure. — Three groups at each level served as learning groups and the fourth as a control group. Group I subjects at each level were given mimeographed sheets on which two problems were presented with one solution for each. The

* An abstract of a dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Education, University of California, 1942.

subjects were instructed to memorize the solutions. They were given no aids or suggestions that might encourage them to look for a principle in the problems, although they were not forbidden to do so. Group II subjects at each level were given similar sheets with the addition of a verbal rule that contained the mathematical principle of the problems: "All lines which form sides of two squares at the same time must be changed to lines which form sides of one square only." They were instructed to study the rule and its application to the problems. Subjects of Group III were given sheets on which non-verbal clues were given that might help them to understand the principle of the problems. In a progressive series of drawings from the problem to the solution, certain squares were shaded in and the lines to be moved were marked. Shading in certain squares changes the appearance of the figure by drawing attention to individual squares and suggests that the squares shaded in must be retained, while others may be destroyed by moving one or two lines. The subjects were told to try to understand how the clues or helps given led to the solution. On all the mimeographed learning sheets the same two problems and the same solutions were presented. The only difference was in the method by which the subjects were instructed to learn to solve the problems. A seven-minute learning period was allowed.

Immediately after the learning period a test of five problems was administered. The control groups took this test without any previous experience with the problems except a short explanation of their nature. The five problems of the test were new problems, but were of the same type as those studied by the learning groups. Any subjects who had seen the problems before and knew the solutions were eliminated. Three minutes were allowed for each problem. A retest of six problems was administered to all groups four weeks after the test. The retest consisted of the two learned problems, two problems repeated from the test, and two new problems. Two minutes were allowed for each problem. After taking the retest, college subjects were asked to mark as "old" any problem they clearly remembered seeing before and as "new" any problem that they were sure they had not seen before. Subjects who said that they had tried to solve the problems in the interval between the test and retest, or who had showed them to others, were eliminated from the experiment.

The results were tabulated by groups at each of the three educational levels. Only perfect solutions were scored as correct. The form of solutions used was tabulated for one problem of the test and for the two learned problems of the retest. In order more exactly to determine the effects of mental age on ability to solve the problems, the subjects were then rearranged into low, average, and high mental age sections regardless of the educational level to which they belonged. The scores were then retabulated by learning groups at the three mental age levels of 15 years MA and lower, 15–1 years MA to 16–9 years MA, and 17 years MA and higher for low, average, and high mental ages respectively.

Findings.—The results are as follows:

1. The mean test scores of all learning groups were significantly higher than those of the control groups at the same level.

2. Differences between the test scores of learning groups were small. At the junior high school and college levels Group II, the subjects given the verbal principle or rule, had a higher mean score than the other learning groups, while at the senior high school level the highest mean score was made by Group I, the subjects who had learned by memorizing. Group III, the group given non-verbal clues, made the lowest mean score at junior high school level, but were only slightly below Group II at the college level. None of the differences between learning groups was fully significant, but the differences between Group II and Group I and between Group III and Group I at the college level approached statistical significance.

3. Every senior high school learning group made a considerably higher mean score than any junior high school group, and every college learning group made a higher mean score than any senior high school group. With only two exceptions these differences are statistically significant.

4. The subjects given the verbal rule made the highest mean score on the retest at each educational level. Group I and Group III followed in order except for the junior high school level, where they were equal. The differences between groups are not statistically significant, but the probability of so

many differences in the same direction occurring by chance is very small.

5. The subjects who memorized the solutions of two problems during the learning period had no advantage over the "rule" group in solving these learned problems on the retest. Group III made the lowest percentage score (percentage of possible solutions) on the learned problems at each educational level.

6. Group II made the highest percentage score on tested problems at each educational level.

7. Group II had a distinct advantage over the other groups in solving new problems. At the junior high school level Groups I and III made lower scores than the control groups.

8. Subjects of Group I were somewhat more likely than subjects of other groups to reproduce the learned solution when presented later with the same problem or one for which the same solution is possible.

9. Group I college subjects who recognized the learned problems as ones they had seen before had a much better chance of solving them than did those who did not recognize them. Group II and Group III subjects were less dependent upon recognition and solved both learned and tested problems equally well whether they recognized them or not.

10. The coefficients of correlation of test scores with intelligence test scores and of retest scores with intelligence test scores ranged from .19 to .47. Five of the seven coefficients that are clearly significant are correlations of Group I scores with intelligence test scores.

11. The coefficients of correlation between test and retest scores ranged from .38 to .73, with the highest coefficients for learning groups being those for Group I subjects.

12. When the data were retabulated by mental age levels, the scores of learning groups were well above those of the corresponding control groups at each level.

13. The subjects of average mental age made approximately twice as high mean percentage scores as did those of low mental age on both the test and retest, regardless of the learning method used. The subjects of high mental age made mean percentage scores approximately three times those of subjects of low mental age.

14. Within each mental age level the differences between the scores of the learning groups were comparatively small. The subjects who were given the verbal rule had a small but definite advantage at each mental age level, particularly on the retest, with the largest margin of advantage at the average mental age level.

Conclusions.—The conclusions are:

1. Within limits set by the mental ability of the learner, memorizing, the use of a simple verbal principle or rule, and the use of non-verbal clues are all reasonably efficient methods of learning.

2. A verbal statement or rule expressing simply the principle of the problems is a more efficient aid to learning than either memorizing the solution of a type problem or providing the learner with non-verbal clues. This is true for learners varying widely in mental ability and from junior high school to college in educational level. The insight gained from this method is more complete than that gained from the use of the other learning methods tested in this study.

3. As compared with the use of a verbal principle, memorizing as a method of learning enables the learner to gain a more limited degree of insight which is most useful in a situation similar to the one in which learning occurred, and of limited usefulness in solving new problems on a delayed test. Gaining insight by memorization seems to be slightly more dependent upon the level of intelligence of the learner than does gaining insight by the other methods used in this investigation.

4. As measured by immediate transfer to new problems, the method of non-verbal clues used in this study is slightly less efficient than providing the learner with the principle in verbal form; as measured by retention it ranks lowest, and as measured by ability to solve new problems in the retest it is not superior to memorizing.

5. The mental ability of the learner appears to be a more important factor in the efficiency of learning than variations in the method of learning.

Bibliography

1. Hartmann, George W. *Educational Psychology*, p. 298. New York: American Book Co., 1941.
2. *The Forty-First Yearbook of the National Society for the Study of Education: Part II, The Psychology of Learning*, pp. 156, 158, 191, 193. Bloomington, Illinois: Public School Publishing Company, 1942.
3. Woodworth, Robert S. *Psychology*, p. 299. Fourth Edition. New York: Henry Holt and Co., 1940.

FACTORS ASSOCIATED WITH THE ACHIEVEMENT OF HIGH SCHOOL PUPILS OF SUPERIOR INTELLIGENCE[1]

JOHN W. MUSSELMAN
Southern Junior–Senior High School
Baltimore, Maryland

THE PROBLEM

Why do intelligent pupils often fail or do mediocre school work? This study of 297 Baltimore high school pupils has been set up to answer this problem. Pupils of superior intelligence are defined as pupils who have an I. Q. of 120 or more, as measured by the Otis Self-Administering and the Henmon–Nelson Tests of Mental Ability.

It would be desirable to study the relation of every possible variable to the discrepancy between promise and performance in pupils of superior intelligence. Time, however, has limited this study to two fields of related variables, personality and home backgrounds.

Educators have long realized that attention and special help should be given to superior children. Du Pont de Nemours stated in 1800 in his "National Education in the United States" that "a single day of an educated man of genius is of more value to the world than the labor of 100,000 men for a year."

Many educators since that time have struggled to help our superior pupils so as to reduce the lag that so often exists between promise and performance.

PREVIOUS STUDIES

Earlier investigators are in general agreement that in individual cases clinical study has frequently resulted in locating factors associated with poor school adjustment and in prescribing effective remedial measures. This is well brought out by such authors as Hollingsworth[2] in her study of gifted children.

Studies of groups of intelligent pupils have not been so successful in bringing to light the causes of maladjustment or in prescribing the remedy for poor school achievement. Factors of heterogeneity are introduced that are not present in the study of an individual case. Although the literature of pupil achievement is often not in agreement, certain hypotheses are well accepted and supported. It is generally agreed that the mean correlation between intelligence and achievement is near .50. Recent evidence indicates that children of foreign born parents achieve better in relation to their intelligence than do children of native born parents. Unhappy family relationships appear to have a detrimental effect on school achievement. When general health is considered, there seems to be considerable evidence that the relationship between health and school success is slight but positive, and that certain types of physical defects are more retarding than others. Good study habits seem to be closely related to sustained effort and superior achievement among superior pupils. Socio-economic status appears not to be important, except where cases of failure and other maladjustment are concerned. The majority of investigators indicate a positive relation between socio-economic status and school achievement. There is considerable disagreement among investigators regarding achievement and parents' education, certain health factors, home discipline, and position in the family or age of siblings. The whole effect of personality upon achievement is in question. These, in general, are the results of previous studies of superior pupils concerning the relation of school achievement to personality and home backgrounds.

RELATION TO OTHER STUDIES

This study varies from other investigations of factors associated with the varied achievement of superior pupils in several ways.

Two relatively new and improved measures of personality have been selected to measure certain aspects of personality. By the use of the Washburne Social-Adjustment Inventory and the Link Inventory of Activities and Interests with the same selected group of pupils of superior intelligence, the writer hopes to

[1] Summary of a dissertation under the same title, submitted as a partial requirement for the Ed.D. degree at The Johns Hopkins University School of Education, 1942.

[2] Leta S. Hollingsworth, *Gifted Children: Their Nature and Nurture*. New York: Macmillan Co., 1926. Pp. xxiv + 374.

establish the existence of new associations between the achievement ratio and personality traits, which might not otherwise have been brought out.

A number of infrequently studied variables believed to be associated with the discrepancy between promise and performance for pupils of superior intelligence have been included in this study of the relation of achievement to home backgrounds. Among these are home discipline, rural versus urban environment, and certain aspects of family and sibling relationships.

Considerable work has been done in investigating the association between achievement and the Sims socio-economic status, study habits, antecedents' country of birth, foreign languages spoken in the home, parents' and pupils' health and physical defects, and parents' education.

For these variables, this study adds a new viewpoint in investigating these relationships by using the achievement ratio.

A relatively recent method of holding intelligence constant has been used, the achievement ratio. This measure of the degree to which a pupil's school performance is approaching his promise has several advantages. By using the A. R. it is possible to include many cases that might have to be dropped, if pairing or matching or some other system of equating for intelligence is used.

By providing a measure of how well performance approaches promise for each pupil, the A. R. allows greater opportunity for dividing a variable to be studied into subcategories and by making it possible to obtain an A. R. for each category, the A. R.'s for the categories may be compared. In addition, the achievement ratio lends itself nicely to statistical calculation.

Care has been taken to make this sample of pupils of superior intelligence as representative of metropolitan high school pupils as possible. The pupils have been drawn from all four senior high school grades of five high schools from different sections of Baltimore, Maryland. These pupils have been so selected as to approach closely a cross section of white metropolitan high school population of Baltimore. Baltimore is itself a representative city of a representative state in the United States.[3]

[3] Howard M. Bell, *Youth Tell Their Story: A Study Conducted for the American Youth Commission*. Washington: American Council on Education, 1938. Pp. 273.

It is hoped that, as a result of these changes in approach to the problem, new data will be uncovered that will make a contribution to the field of research on superior pupils.

Selection of Cases

The 297 cases included in this study were selected from five white public secondary schools from different sections of the city of Baltimore. The combined enrollment of these schools totaled approximately 11,000. Pupils were selected from the various schools on a voluntary basis, with their parents' consent. Only those pupils from grades 9–12 whose Otis intelligence quotients as recorded by the schools were 120 or above were included in the preliminary invitation. If they scored 120 I. Q. or above in another form of the Otis, and in the Henmon–Nelson Test of Mental Ability, Form A, they were retained in the study. Eighty per cent of those invited accepted the invitation, qualified by passing the Otis and Henmon–Nelson Intelligence tests, and completed the necessary tests and questionnaires. The 143 boys and 154 girls finally retained in the study represented every section of Baltimore.

Collection of Data

Besides the two group intelligence tests administered to every pupil, the following additional tests and questionnaires were given to the selected pupils:

Achievement

Either the Myers–Ruch High School Progress Test, Form Am for Senior High Schools; or the Sones–Harry High School Achievement Test, Form A.

Personality

Washburne Social-Adjustment Inventory (Thaspic edition).

H. C. Link's Inventory of Activities and Interests (P.Q. test).

Home Backgrounds

Sims Score Card for Socio-Economic Status, Form C.

Wrenn and McKeown Study Habits Inventory.

Physical Examination Questionnaire filled out by the school doctor.

A general questionnaire of seventy items supplying information on home backgrounds.

In addition, the scholarship and packet records of the schools were used to supply information on subject marks, conduct and effort marks, attendance records, various subject matter and intelligence test scores, early physical data on entrance to first grade, any important information concerning the pupil's school difficulties or problems, pupil's aim for life, vocational choice, interests, scholarship record from the first grade to date, and any other matters of special interest to the principal or vocational counselor.

Treatment of the Data: the Achievement Ratio

By combining the measures of intelligence and the measure of achievement, a unique measure for each pupil was obtained, the achievement ratio. The achievement ratio is a measure of the discrepancy between promise and performance. It may be used to determine whether a pupil's school progress or achievement is poorer than, equal to, or better than his general progress as indicated by an intelligence test. It is possible for a pupil to be progressing scholastically in school better than we would expect one of his intelligence to achieve. Unfortunately, the reverse is more often true, particularly in the case of pupils of superior intelligence. The achievement ratio used here is, then, a measure of how well a pupil's performance in school, as measured by a standardized achievement test, approaches (is less than, equal to, or better than) his promise, as measured by an intelligence test.

The achievement ratio is arrived at in this manner. The pupil's achievement test score is changed to a standard score; the pupil's intelligence test score is changed to a standard score; and then the intelligence standard score is subtracted from the achievement standard score. The result is the achievement ratio. Where the achievement ratio is positive, it is an indication that the pupil is achieving better than might be expected of a pupil having his intelligence. Where the achievement ratio is negative, it is assumed the pupil's performance is below his promise.

The relation of the pupils' achievement ratios to a number of variables in the general fields of personality and home backgrounds was determined. The Washburne and Link personality subtest and total scores served as one group of variables. Twenty-two variables in the field of home backgrounds include those covering parent-pupil relationships, the pupil's antecedents' birthplace, parent and pupil health, sibling relationships, home study conditions, and general home environment.

A graph showing the relation of achievement ratio to each variable and its subclassifications was made. Separate frequency distributions and graphs were made for the boys and girls, as well as the total group. However, these graphic illustrations are not included here.

Analysis of Variance

The method used for calculating the relationship between a certain variable and the achievement ratios was analysis of variance. By using the "F" test, it was possible to determine whether a significant relationship existed between any particular variable and the achievement ratio. As Lindquist[4] says: "It must be remembered when using the F test that the degree of relationship is not given, and that while a relationship may be consistently present, it may at the same time be *small*."

Data on Achievement

Data on achievement test scores for the group indicate that the superior pupils are considerably above average for achievement, and that the boys excel the girls in mean achievement test scores.

The data on mental ages indicate for the boys a slight superiority over the girls.

The data on I. Q. give the boys superiority over the girls in mean Otis Intelligence Quotients and the Henmon–Nelson Intelligence Quotients. The mean I. Q. for the total group on the Otis is 123.6 and on the Henmon–Nelson is 126.6.

Data on achievement ratios show for the boys definite superiority over the girls and indicate that the achievement of both boys and girls in high school subjects is considerably lower than might be expected, using their intelligence scores as measures of promise.

The data on general scholarship averages indicate that these superior pupils are doing good work on the average and that the boys are doing slightly better than the girls.

The data on conduct and effort marks indicate that the pupils of superior intelligence

[4] E. F. Lindquist, *Statistical Analysis in Educational Research*, p. 16. Boston: Houghton Mifflin Co., 1940.

receive relatively few unsatisfactory conduct and effort marks in high school and that the boys receive more unsatisfactory marks in conduct and effort than the girls.

The data on "semesters failed" show that the boys fail more semesters in high school than do the girls, but that either a very small number of superior pupils fail semesters in high school or, if they do fail, they do not continue in high school.

The data on "semesters accelerated" in school show that the girls skipped more semesters than boys throughout their school career, and that the average number of semesters skipped is just a trifle over one for the pupils studied.

There is nothing especially startling about these data except that wherever the boys are compared with the girls, the girls suffer by comparison except in the matters of semesters skipped, semesters failed, and conduct and effort marks. The boys excel the girls in mean achievement test scores, mean mental ages, mean intelligence quotients, mean achievement ratios, and mean scholarship averages.

The Relation of Achievement Ratio to Personality

Introduction

Two personality inventories were administered to every pupil participating in the study. The first of these inventories, the Washburne Social-Adjustment Inventory, Thaspic Edition, is made up of seven subtests. The second inventory, the Link Inventory of Activities and Interest (Personality Quotient), is made up of four subtests and an overall score from which is obtained the personality quotient.

A. The Relation of Achievement Ratio to Truthfulness

A screening test of truthfulness was given, composed of questions scattered throughout the inventory, designed to reveal carelessness or misrepresentation in replies. These questions were of such a nature that apparently desirable answers indicate inaccuracy or lying. The great majority of pupils made low scores on this subject, indicating excellent adjustment for truthfulness. The "F" test, giving a value of 10.531, indicates that a significant relationship existed between achievement ratios and truthfulness scores. Reference to the graphic illustration of the relationship, not included here, revealed that the most truthful pupils had the highest mean achievement ratios, while the untruthful pupils had the lowest mean achievement ratios. This relationship held with little variation for both boys and girls.

B. The Relation of Achievement Ratio to Happiness

The great majority of superior pupils, 233 out of 297, were normal or better in their adjustment for happiness, according to the norms set up by Washburne. The F for the total group was 7.56, indicating 99 chances out of a hundred that a significant relation existed between achievement ratio and happiness. In achievement ratio, the well adjusted happy pupils ranked above the pupils who were low normal in adjustment for happiness, but definitely below the pupils who were maladjusted, as indicated by the happiness scores. For boys, girls, and the total group, the well adjusted and poorly adjusted pupils excelled the low normal happiness cases in mean achievement ratio. It may be that up to a certain point, the more unhappy you are, the less your achievement ratio will be, due to a poor adjustment to your surroundings; but, beyond this point, an increase in unhappiness may cause you to spend more time in individual activities not associated with a social group, which in turn may result in an increased achievement ratio.

C. The Relation of Achievement Ratio to Alienation

Alienation as defined in this inventory includes a sense of social membership, similarity, psychological security, and emotional stability in social situations. Of the 297 pupils, 222 were normal or above on this trait. The "F" test yielded a value of 6.247 for the total group, which was significant at the 1 per cent level. The relation between achievement ratio and alienation was uniformly negative. As achievement ratio increased, alienation adjustment decreased. There was some fluctuation in the relationships for boys and girls, but the graphic illustration for the total group maintained a consistent slope.

D. The Relation of Achievement Ratio to Sympathy

The inventory defines sympathy as empathetic, non-negative responsiveness to people. There were 251 pupils normal or better for this trait. The F was high, 13.38, indicating a significant relationship that turned out to be negative. This relationship was characterized by considerable fluctuation, but it was evident that the high achievement ratios were associated with the poorer adjustments for sympathy. The girls were markedly more sympathetic than the boys.

E. Relation of Achievement Ratio to Purpose

Purpose is defined by Washburne as desire directed toward a goal involving plan, evaluation, selection, and effort. Of the 297 superior pupils, 218 were normal or better in this trait according to norms set up by Washburne. The F test yielded a value of 11.974 which, with the given degrees of freedom, indicated less than one chance in a hundred that no relationship existed between achievement ratio and purpose. If the fluctuation at the ends of the curve where cases are few is disregarded, it is observed that the relation is curvilinear, assuming the form of a flat bell shaped curve. High achievement ratio is associated with normal or intermediate adjustment for purpose.

F. Relation of Achievement Ratio to Impulse-Judgment

Impulse-judgment is the ability to judge well between conflicting impulses. There were 259 pupils rated as normal or better on this trait. The F test of 6.689, coupled with the graphic illustration, indicated a significant, inverted flat bell-shaped curvilinear relation between achievement ratio and impulse judgment. Low achievement ratio is associated with intermediate impulse-judgment adjustment. The girls excel the boys in adjustment on this trait, but for boys the relation between excellent impulse-judgment and high achievement ratio is markedly greater.

G. Relation of Achievement Ratio to Control

Satisfactory control adjustment includes a sense of self-regulation and the ability to make and execute plans. The great majority of pupils, 261, have low normal or better adjustments on this trait. The F value of 6.002 and the graphic illustration of the relationship indicate a significant relation that is curvilinear, of the inverted bell type. Low achievement ratio is associated with intermediate or average adjustment for control. High achievement ratio is associated with both poor and excellent adjustment for control.

H. Relation of Achievement Ratio to Total Social Adjustment

The Washburne total score is a composite score made up of a total of all the subtest scores. It is a measure of general social adjustment. For the 297 pupils tested, 34 were definitely superior in their adjustment, as measured by the Washburne S-A Inventory, 85 were high normal in adjustment, 55 were low normal in adjustment, 67 were borderline cases, and 10 pupils were definitely maladjusted. In general, girls were better adjusted than boys. The F test, coupled with the graphic illustration, indicated that a significant relation existed between achievement ratio and general social adjustment, which was curvilinear of the inverted bell type. The curve showing this relation of achievement ratio to Washburne social adjustment begins at a high level, drops in the middle, and rises again to a level nearly as high. Low achievement ratio is associated with average social adjustment. Both poor social adjustment and excellent adjustment are associated with high achievement ratio, with pupils of poor social adjustment having slightly higher achievement ratios. With little or no variation, this relation holds for both boys and girls, as well as for the total group of superior pupils studied.

It is interesting to note the fact that the girls as a group seem to be a trifle better adjusted than the boys, as measured by the Washburne Social-Adjustment Inventory. The chief part of the difference is probably brought about by the better adjustment of the girls on such traits as truthfulness, happiness, alienation, sympathy, and impulse-judgment subtest scores. The girls' scores indicate only slightly better adjustment than the boys, except in the trait of sympathy where the adjustment of the girls is definitely superior to that of the boys. For the traits of purpose and control, the boys excel the girls in adjustment.

Perhaps this significant relationship between Washburne S-A Inventory total scores and achievement ratio may be explained in the following way. If we assume that pupils who are not well adjusted socially, as measured by the Washburne S-A Inventory, do not get along well with their companions, it is quite probable that they associate and mingle less with these companions. These poorly adjusted people probably spend less time in social activities and more time on interests and tasks involving individual attention. Perhaps, because of this comparative seclusion brought about by fewer social contacts with their companions, they have more time for and turn their interests toward school accomplishment. As a result of this increased attention to school tasks, they are likely to have high achievement ratios.

On the other hand, using this same line of argument, it would seem that those pupils who have more social contacts and are better socially adjusted with their companions will probably spend less time on their school work and be less interested in school accomplishment, and consequently have lower achievement ratios. This is apparently what happens up to a certain point, but beyond that point, the better a pupil is adjusted, the higher his achievement ratio becomes. This may be explained by the assumption that excellent social adjustment carries with it advantages that counterbalance relative seclusion from a social group, which is perhaps the greatest advantage of the poor adjustment group of pupils. It may be that excellently adjusted pupils, by the very reason of their increased social contacts, as well as more efficient use of their time, are able to increase their achievement ratios to a point where they excel the achievement ratios of the intermediately adjusted group.

Link Personality Quotient Inventory

The Link Inventory of Activities and Interests or P. Q. test is made up of four subtests from which is obtained an overall personality score that may be reduced to a Personality Quotient. The personality quotient is intended to be for personality what the intelligence quotient is for intelligence.

A. Relation of Achievement Ratio to Social Initiative

The first subdivision of the Link Inventory, social initiative, includes habits of taking the initiative in meeting and dealing with people. No significant relation was found to exist between achievement ratio and social initiative. The graphic illustration of the relationship, however, indicated that a negative association existed.

B. Relation of Achievement Ratio to Self-Determination

The scale for this trait consists of items representing habits of doing things that are considered desirable, even though they may be unpleasant. The distribution of scores on this trait is close to normal. The mode and mean of the distribution for the total group fall in the mid-interval of the distribution of self-determination scores. The F value of 2,994, coupled with a graphic illustration, indicates that a negative significant relation of small degree exists between achievement ratio and self-determination. High scores on the self-determination scale, indicating excellent adjustment, are associated with low achievement ratio. It is also true that at the other end of the scale poor self-determination adjustment is associated with high achievement ratio. If only the intermediate group of pupils who have intermediate self-determination scores and achievement ratios is considered, no relation at all is observed.

C. Relation of Achievement Ratio to Economic Self-Determination

This scale of the Link inventory[5] measures the ability to earn one's living; it consists of various habits of earning money. The frequency distribution of economic self-determination scores makes up a normal distribution, with the mean and mode falling in the mid-interval. The F value of 5.326 indicates that there is less than one chance in a hundred that the null hypothesis is true. Reference to the graphic illustration of the relationship for the total group (not included here) shows that the relation between achievement ratio and economic self-determination is fluctuating, negative, and small.

The three groups of pupils having the highest mean achievement ratios fall at three widely separated intervals in the curve. The first group of pupils has very low E. S. D.

[5] Henry C. Link and Others, *Manual for the P.Q. or Personality Quotient Test: Inventory of Activities and Interests.* New York: Psychological Corporation, 1938 (revised). Pp. 14.

scores, indicating poor adjustment for this trait. The second group of pupils having a comparatively high mean achievement ratio falls in the middle of the curve for the total group, showing average adjustment on E. S. D. The third group of high achievement ratio pupils falls near the upper end of the E. S. D. curve, indicating good adjustment for E. S. D. General inspection of the curve for the total group reveals considerable fluctuation, particularly at the ends of the distribution where the number of cases is small. If the ends of the distribution of achievement ratios were to be removed, the distribution would be of the flat bell-shaped curve type.

When the distributions for the boys and girls are compared, it is observed that both are characterized by considerable fluctuation. The boys who have the highest mean achievement ratios fall at the middle of the curve. The graphic illustration of the relationship between the achievement ratios for the girls and E. S. D. scores shows considerably less fluctuation than that for either the boys or the total group. The girls with the lowest achievement ratios have definitely the highest E. S. D. scores, while the girls with the highest achievement ratios have the lowest E. S. D. scores. The negative relationship between E. S. D. and achievement ratio is consequently more marked for the girls than it is for either the boys or the total group.

The small significant negative relationship that exists between E. S. D. and achievement ratio may be accounted for in several ways. Pupils who are economically self-reliant or who rank high in ability to take care of themselves economically while attending school are more likely to be spending a considerable part of their time at certain after school jobs, which will reduce the amount of time they can spend on their studies; consequently, their achievement in school will not come as close to fulfilling their promise, as might otherwise be expected. The degree of the relationship is small and, in the opinion of the writer, merely presents evidence that no positive relationship exists between the achievement ratio of pupils of superior intelligence and what Link calls economic self-determination. This finding seems to be in agreement with the findings of others with regard to matters of vocational choice, etc.

It may be, too, that the association of low achievement ratios with good adjustment, as far as economic self-determination is concerned, may result from other factors influencing this relationship. High rating for E. S. D. may be associated with need or intention to leave school early. This in turn may influence the achievement ratio to some considerable degree. Other factors may influence the relationship indirectly in just such a manner.

D. Relation of Achievement Ratio to Adjustment to the Opposite Sex

This scale assumes that proper sex adjustment depends upon habits of action formed early in life toward members of the opposite sex. The distribution of sex adjustment scores is skewed to the left as a result of the preponderance of scores falling at the well adjusted end of the distribution. Unquestionably, according to Link, the great majority of these pupils of superior intelligence are well adjusted toward the opposite sex. It should be noted that, of the forty-one cases in the total group showing the best sex adjustment, all were girls. It further appears that the distribution of boys' scores for sex adjustment is normal, but the distribution of girls' scores for sex adjustment is quite skewed to the left.

When the F test is applied to the distribution, a value of 13.552 is obtained. With the degrees of freedom involved, the chances of a relation between achievement ratio and adjustment to the opposite sex not existing are less than one in a hundred. Reference to the graphic illustration of the relationship reveals that the significant relation between achievement ratio and adjustment to the opposite sex is definitely negative. Pupils having high achievement ratios are likely to be poorly adjusted to the opposite sex, while pupils having low achievement ratios are likely to be well adjusted to the opposite sex. This relationship holds for both boys and girls, as well as for the total group.

There is some fluctuation in the linear relationship that seems to exist between achievement ratio and adjustment to the opposite sex scores, and this fluctuation increases for both the boys and girls.

The negative association between achievement ratio and adjustment to the opposite sex, which holds for the boys, girls, and total group, may be accounted for in a number of ways. The pupils who acquired, probably

through a number of personal contacts with the opposite sex, an excellent adjustment to the opposite sex, probably spend less time and are less interested in school work. Consequently, they achieve considerably below what might be expected of one of their intelligence.

The fact that the girls are so much better adjusted to the opposite sex than the boys might in a large measure account for the fact that the mean achievement ratio for the girls is considerably lower than the mean achievement ratio for the boys. On the other hand, pupils who do not do so well in school or whose adjustment comes further from their promise than might ordinarily be expected may turn to the opposite sex for an outlet for their energies.

E. The Relation of Achievement Ratio to Personality Quotient

This scale includes all the items on the Link inventory. The scale measures the possession of habits that interest and serve other people. The personality quotient may be interpreted as the intelligence quotient, a score of 100 being average. On this inventory 150 pupils received P.Q.'s of less than 98, while 103 pupils received P.Q.'s of over 105. The mode for this particular test fell at the 98 to 105 interval, where there were 55 cases. Judging from these results, it appears that this superior group of pupils has just average personality quotients. The distribution of personality quotients is just about normal for the girls, as well as for the total group. The distribution of boys' P.Q.'s was skewed a bit to the right, with a lower mean P.Q. than that for the girls.

When the F test is applied to the total group, a value of 8.941 is obtained, which is significant at the one per cent level. Observation of the graphic relationship indicates that the relation is negative to a small degree, but with a significant fluctuation. When the group as a whole is studied, it is observed that the mean achievement ratio for the mid P.Q. interval (98 to 105) is lower than for any other category selected. It appears that average personality quotient is associated with low achievement ratio. It is also true that both high and low personality quotients are associated with high achievement ratio. The pupils having the lowest personality quotients have the highest achievement ratios, and the pupils having the highest personality quotients have slightly lower achievement ratios. This association is true for boys, girls, and the total group. These findings agree essentially with those obtained from the Washburne S-A Inventory, and may be accounted for in the same way.

Relation of Achievement Ratio to Home Backgrounds

A. Relation of Achievement Ratio to Socio-Economic Status

The Sims Score Card for Socio-Economic Status is a simple, convenient, and objective device for determining the cultural, social, and economic backgrounds furnished by the homes of school children. Scores on the questionnaire may be arranged in ten socio-economic status classes set up by Sims, ranging from indeterminately low to indeterminately high. The great majority of pupils fell in the medium to highest classes. Boys and girls differed little in this respect. This group of superior pupils was definitely above average in socio-economic status.

The variance ratio F for the relation of achievement ratio to socio-economic status was 6.132. The degrees of freedom were sufficient to make this significant at the 1 per cent level. There is less than one chance in a hundred that no relation exists. The graphic illustration of the relationship indicates that it is linear and positive. High achievement ratio is associated with high socio-economic status for pupils of superior intelligence.

With the exception of slight fluctuations at the end of the curve, this significant positive linear relationship between achievement ratio and socio-economic status appears to exist for both boys and girls, as well as the total group. The boys at the lower end of the S. E. S. distribution do not have the very lowest achievement ratios for the boys, and the girls at the upper end of the S. E. S. distribution do not have the very highest achievement ratios for the girls; with these two exceptions, however, the A. R.–S. E. S. curve is uniform, indicating a definite significant positive linear relationship between achievement ratio and socio-economic status for boys, girls, and the total group.

This relationship between achievement ratio and S. E. S. may be accounted for in a number of ways. In general, a better social and economic status possibly carries with it

advantages that help the pupil in making a higher achievement ratio. Such things as possession of more books in the home to refer to, more rooms in the house allowing more privacy for study, better educated parents, etc., may all aid the pupil in making a higher achievement ratio. Apparently boys of low socio-economic status have less trouble than girls in overcoming this disadvantage and, consequently, make better achievement ratios. Apparently, too, a very high achievement ratio is a disadvantage for girls, but not for boys. Girls of high socio-economic status possibly have relatively more outside interests, with the result that their school work suffers, and their achievement ratio falls.

B. Relation of Achievement Ratio to Foreign Languages Spoken in the Home

Every pupil included in the study was asked to list the number of languages, including English, spoken in his home. Twelve foreign languages were spoken in the homes of 83 pupils. The remainder of the group, 214, lived in homes in which only English was spoken. German and Yiddish were the two foreign languages most frequently spoken in the homes of these superior children. In sixteen homes Lithuanian, French, Polish, Russian, Hungarian, Greek, Rumanian, Bohemian, or Spanish was spoken. The F test of the significance of the relationship between pupils in whose homes a foreign language was spoken and pupils from purely English-speaking homes yielded a value of 15.267. This indicated that there was less than one chance in a hundred of the difference being due to chance. The graphic illustration of the relationship showed that people from homes in which foreign languages were spoken excelled those from only English speaking homes. Ranked according to languages spoken in the home, the miscellaneous group of 16 cases was first in achievement ratio, the Yiddish group of 51 cases was next, the German group of seventeen cases third, and the purely English-speaking group of 214 cases last.

This association between high achievement ratio and coming from a foreign-language-speaking home held good for the boys, as well as the total group. The girls, however, from homes in which no foreign language was spoken excelled those girls from both German- and Yiddish-speaking homes in mean achievement ratio. These girls were excelled only by the miscellaneous group. Apparently the significant relationship existing between achievement ratio and the presence of foreign languages spoken in the home is due largely to the influence of the boys.

If a foreign language is spoken in the home, it is reasonable evidence that some members of the family have not been in America during their entire lifetimes, certainly not long enough to forget how to speak their native tongues. In some cases the fact that the foreign languages are spoken in the home may be evidence of superior education and knowledge of languages not native to the family, but this is probably the exception rather than the rule.

If, then, we assume that the speaking of foreign languages in the home is a good criterion that people in the homes have been immigrants not too long ago, how can we account for the fact that pupils coming from homes in which foreign languages are spoken have a higher mean achievement ratio than "native" pupils? What is the difference between the boys and girls that accounts for the lower mean achievement ratio for the girls who come from homes in which foreign languages are spoken? Perhaps the boys from foreign-language-speaking families feel that they have a handicap and work harder to overcome it. The boys come in greater contact with the outside American world and possibly are more influenced by it than girls may be. These are possible explanations for the differences found.

C. The Relation of Achievement Ratio to Antecedents' Country of Birth

Every pupil in the study was asked to find out and list the country of birth of his mother, father, and four grandparents. The achievement ratios of pupils of native born antecedents were compared with those of foreign born antecedents.

The number of American born grandparents of any one pupil was selected as the first variable. The great majority of pupils (125) had four American born grandparents, 17 pupils had 3 American grandparents, 29 pupils had two, 6 pupils had one, and 96 pupils had no American born grandparents. Twenty-four pupils did not know where all their grandparents were born. The F test applied to these data revealed a significant

relation between achievement ratio and number of American born grandparents. The graphic illustration showed that the fewer American grandparents a pupil has, the higher his achievement ratio is likely to be. There is some fluctuation, but the general trend is unmistakable for boys as well as girls.

Being of recent foreign extraction apparently supplies a stimulus to achievement. This finding is similar to that discovered for the relation between achievement ratio and foreign languages spoken in the home, and is in general agreement with the results of Desing[6] and other investigators.

The relation of achievement ratio to individual parents' and grandparents' country of birth was next treated. Only those countries in which ten or more parents or grandparents were born were included in separate categories. Countries in which less than ten antecedents of any one kind were born were combined into a miscellaneous category. Table I lists the pupils' mean achievement ratio ranks for each type of antecedent.

It will be noted here that pupils whose antecedents were born in the United States rank lowest in mean achievement ratio for each type of pupil antecedent considered. Apparently, pupils whose antecedents are native born have lower mean achievement ratios than pupils whose antecedents are foreign born. The F test bears out this hypothesis in every case. This is still another evidence that superior pupils of foreign extraction are likely to excel superior pupils of domestic antecedents in achievement ratio.

This consistent relation may be accounted for in several ways. Pupils whose antecedents are foreign born are more likely to expend more effort to overcome language and cultural differences and consequently have higher achievement ratios than pupils whose antecedents are native born. Another explanation of this relationship between achievement ratio and the foreign born antecedents of superior pupils may be that these pupils having foreign born antecedents are more likely to appreciate the opportunities offered in America where the opportunity is often greater than that enjoyed in their parents' or grandparents' native lands. Because this is so, in many cases the appreciation of the opportunities offered in America may be instilled in the pupil by his foreign born parent or grandparent.

D. Relation of Achievement Ratio to Parents' Education

Data on the fathers' education for the pupils studied revealed that there is a significant relation between achievement ratio and the number of years of schooling the pupils' fathers have. The F value of 20.496 indicated that there is less than one chance in a hundred that no such relation exists. The graphic illustration shows that the more formal education a pupil's father has, the higher the pupil's achievement ratio is likely to be. It appeared, however, that the achievement ratios of boys were less influenced by the father's education than were the achievement ratios of the girls.

The formal education of the mothers of these superior pupils ranged from attendance at elementary school to graduation from college. The median mother attended high school, but did not graduate. The F of 12.248 for the total group was significant at the one per cent level. The graphic relationship indicates that an increased achievement ratio is associated with increased mothers' education. Again the achievement ratios of

[6] Minerva F. Desing, *The Relation of Pupil Achievement Gains to Certain Personal and Environmental Elements*, p. 144. Philadelphia: University of Pennsylvania, 1940.

TABLE I

MEAN ACHIEVEMENT RATIO RANK OF PUPILS OF SIX TYPES OF ANTECEDENTS SEPARATED BY COUNTRY OF BIRTH

Type of Pupil Antecedent	Country of Birth				
	Russia	Poland	Germany	Misc.	U. S.
Father	2	3	1	4	5
Mother	2	1	------	3	4
Father's father	3	1	2	4	5
Father's mother	3	1	2	4	5
Mother's father	3	1	2	4	5
Mother's mother	2	1	3	4	5

[6] Minerva F. Desing, *The Relation of Pupil Achievement Gains to Certain Personal and Environmental Elements*, p. 144. Philadelphia: University of Pennsylvania, 1940.

the girls are more highly influenced by the mother's education than are the boys. With no exception, high mean achievement ratios for the girls are associated with increased education of the girls' mothers. There is no significant association between the achievement ratios of the boys and the mother's education. It is evident that the relationship that exists is due almost entirely to the girls.

Perhaps a pupil's parents who are better educated are better able to help him with his school work and this may result in a higher achievement ratio for the pupil. Pupils are more likely to respect educated parents and consequently work harder to be like them. Parents who have a good education very often appreciate its value and are better able to help and encourage their offspring to work harder in school so that they too may be well educated. Any or all of these reasons may account for the positive significant relation existing between achievement ratio and parents' education. It does seem, however, that parents exert a much greater influence over girls than boys in this respect.

E. Relation of Achievement Ratio to Number of Siblings

The distribution of pupils according to the number of siblings they had was skewed to the right because of the large number of only children and pupils who had only one brother or sister. There is less than one chance in a hundred that a relation does not exist. Reference to the graphic illustration reveals that the smaller the number of siblings the higher the achievement ratio. Children with no siblings rank in achievement ratio just below pupils who have only one sibling, but both of the groups of pupils excel all others in mean achievement ratio. It was very interesting to note that girls with one brother made higher mean achievement ratios than girls with one sister; while boys with one sister had better achievement ratios than boys with one brother. According to this study, pupils having one sibling of the opposite sex are most likely to have the highest achievement ratios. Peculiarly enough, pupils who have two siblings only of the opposite sex have the lowest mean achievement ratios.

F. Relation of Achievement Ratio to Sibling's Position in the Family

The pupils were divided into four categories: the youngest, the intermediate, the oldest pupils, and the only children. There were fewer intermediate children than in any of the other three categories. There were 95 pupils the youngest in their family, 54 neither youngest nor oldest, 88 oldest, and 60 "only" children. The F test coupled with the graphic illustration revealed that a significant relation existed between achievement ratio and sibling's position in the family and that the oldest children made the highest mean achievement ratio. The only children and the youngest children ranked next in mean achievement ratio, while the intermediate siblings were lowest.

The relationship between the achievement ratio and the sibling's position in the family may be accounted for in several ways. The oldest siblings probably have to approach their problems with the least assistance from their younger brothers and sisters and consequently develop habits of action that have a beneficial influence upon their achievement ratio. The youngest children ranked below the oldest children and above the remaining children. Perhaps this is because they receive so much attention, encouragement, and help from their parents and their other siblings that their achievement ratio is beneficially affected. The children of intermediate age are probably the most neglected group, probably having the novelty of the tasks taken off by their older siblings, but getting less attention than the youngest siblings, and consequently having lower achievement ratios than either of the other two groups.

G. Relation of Achievement Ratio to Environment

The measure of environment selected was determined by noting which of four groups the pupil had belonged to as a resident the greater part of his life. The pupils were separated according to whether they had lived the greater part of their life in Baltimore, in any city over 10,000 population other than Baltimore, any town, and pupils living in rural districts. The great majority of the pupils (265) had lived in Baltimore most of their lives. The remaining pupils were lumped in one category and the F ratio calculated. The F of 14.917 indicated that the relationship was significant at the one percent level. The mean achievement ratio for pupils from Baltimore was significantly higher than that for the remaining pupils. Only a small num-

ber of the pupils studied had spent the greater part of their lives in rural communities, but this group ranked lowest in mean achievement ratio.

When the girls alone were considered, it was observed that the mean achievement ratio for native Baltimoreans was much higher than that of all the others. This was not the case when the boys alone were considered. In fact, there was no difference at all between native Baltimore boys and the others in mean achievement ratio. This significant relationship between achievement ratio and this aspect of environment for girls, and not for boys, may be accounted for in this way. The increased advantages of a metropolitan environment have a much greater effect on girls than upon boys. Consequently, their achievement ratios are more highly affected.

H. Relation of Achievement Ratio to Family Residence

By family residence is meant the family or adult group with which the pupil is living. The pupil may be living with both parents, either parent, or neither parent. The great majority of the pupils (251) live with both parents in intact families. The graphic illustration of this relationship indicates clearly that superior pupils coming from broken families had higher achievement ratios. The F value of 4.998 indicated that the association between high achievement ratio and broken families was significant. This relationship was more marked for boys than for girls. This appears to be another case of a handicap offering the incentive to greater achievement. This may be due to the fact that in most separated families the pupils live with their mothers, and the mothers had to work to maintain the home and send the children to school. In the case of a boy attending school, this is probably an incentive to work and learn so he may relieve his mother as soon as possible. Girls apparently are not so highly stimulated by such a situation, possibly because of the increased housework falling to their lot.

I. Relation of Achievement Ratio to Pupil's Health

The school doctors gave each pupil a thorough physical examination. Only fourteen of the 297 pupils were considered to have poor health. No significant relation was found between the doctor's estimate of the pupil's health and achievement ratio.

The number of childhood diseases was selected as another measure of pupil health. There were 233 pupils who had contracted three or more childhood diseases such as measles and whooping cough. It was found that childhood diseases did not differentiate between the pupils as far as achievement ratio was concerned. Any effect that may have existed as a result of such diseases probably disappeared as time elapsed.

The number of physical defects in sight, hearing, teeth, weight, etc., was selected next for study. There were 194 pupils who had no defects of any kind. Again it was found that presence of physical defects had no significant relation to achievement ratio. Those pupils with sight, hearing, weight, and certain miscellaneous defects had higher achievement ratios than the pupils with no defects, but the difference was too small to be significant.

J. Relation of Achievement Ratio to Parents' Health

The great majority (203) of the pupils had parents whose health was considered good. The F test indicated that a significant relationship existed between the parents' health and the pupil's achievement ratio. Pupils whose parents' health was poor had significantly higher achievement ratios than pupils whose parents' health was good. This was true for boys as well as girls. It was surprising to find that the health of the parents had more effect on achievement ratio than the health of the pupils themselves.

The relationship between high achievement ratio and parents' poor health may be accounted for in this way. Those pupils with parents whose health is decidedly poor probably are stimulated by necessity to work as hard as they can in and out of school, in order to take over the burdens of their parents. Then, too, these pupils whose parents have poor health may be given a greater appreciation of the advantages of an education than pupils who do not have this handicap. In all probability, however, this is merely another instance or indication that possession by superior pupils of a handicap tends to make the possessor of that handicap work harder to

overcome it than he ordinarily would; and, in so working to overcome the handicap, his mean achievement ratio would be increased.

K. Relation of Achievement Ratio to Home Discipline

The data on this relationship were acquired from the questionnaire in which the pupils were asked which methods were used at home to discipline them. It was discovered that the majority (121) of the pupils' parents used the method of depriving the pupils of privileges. Sixty-seven parents lectured the pupils, and twenty-three parents used a combination of the first two methods. Ten parents used whipping, twenty-eight did nothing as a method of discipline, and the remainder employed miscellaneous forms of discipline.

The F test indicated that the relationship was significant. There were less than five chances in a hundred that no relationship existed between the various categories under home discipline and achievement ratio.

The graphic illustration indicates that the pupils who were deprived of privileges by their parents as a method of punishment had a lower mean achievement ratio than the pupils who were lectured or punished in any other way. Those pupils who received no punishment at all were ranked next in mean achievement ratio. Apparently, low achievement ratio is associated with being deprived of privileges, and high achievement ratio is associated with whipping and lecturing by parents. The same situation was true in the case of boys, girls, and the total group.

Apparently, a superior achievement ratio is enjoyed by those pupils who are whipped or lectured, because the stimulation of a disagreeable punishment is greater. Depriving of privileges apparently has a greater effect on achievement ratio than any other method of punishment used by parents. It probably has a tendency to make the pupil sulk and want to do nothing at all, which naturally would have an adverse effect upon achievement ratio.

L. Relation of Achievement Ratio to Parents' Working

The achievement ratios for the pupils were compared under these headings: pupils whose fathers only were working, pupils whose mothers only were working, pupils whose mothers and fathers both were working, pupils neither of whose parents nor any one else in the family was working, pupils who had no parents working but relatives working, and pupils who had both parents and relatives working.

The F test for this relationship indicated lack of significance.

For all of the above listed classifications from homes in which no one was working to homes in which parents and one or more relatives were working, there was little difference in the mean achievement ratios of the groups. There is no relationship between the number of people in the home working and the achievement ratio for the group as a whole, as well as for boys and girls.

Apparently, whether everybody or nobody in the family, broken or intact, worked had little effect on the achievement ratio of the pupils studied.

M. Relation of Achievement Ratio to Study Habits

The measure of study habits, the Wrenn[7] Study Habits Inventory, is a list of statements of situations, habits, and conditions that may affect the use of study time, and consequent success in school work and study. According to Wrenn, a negative score on the inventory indicates poor study habits, while a positive score indicates good study habits.

The F test for the distribution yielded a value of 14.937, which was significant at the one per cent level. Evidently a relationship exists between the Wrenn study habits scores and achievement ratio.

Reference to the graphic illustration of the relationship indicates that this relationship is positive. Very low achievement ratios are associated with very poor Wrenn study habits scores, and very high achievement ratios are associated with very good Wrenn study habits scores. The great majority of the pupils who fall between the two extremes of high and low study habits scores varied little in mean achievement ratio. This relationship was true for the boys and girls, as well as the total group. The relationship between study habits scores and achievement ratio can be accounted for very easily.

It was something of a surprise to find that the relationship between study habits scores

[7] C. Gilbert Wrenn and R. B. McKeown, *Study Habits Inventory*, p. 1. Stanford: Stanford University Press, 1934 (revised).

and achievement ratio was not more marked than it was. One would naturally expect that those pupils who have the best study habits would be more likely to make higher achievement ratios. It is only natural to expect that such a relationship would exist, since the two variables seem to be so closely related. There is nothing in this study habits inventory to indicate how much time a pupil spends on his lessons. It might be possible for a pupil with poor study habits, but who spends a great deal of time on his school work, actually to achieve more than pupils who have good study habits but spend little time on their school work. This may account for the fact that the relationship between study habits scores and achievement ratio is not more marked than it is.

Conclusions

1. There is a considerable lag between promise and performance in school, generally for the pupil of superior intelligence. The majority of superior pupils are not achieving in school nearly as much as they are able to accomplish.

2. The lag between promise and performance for boys of superior intelligence is less than that for girls, i.e., the achievement ratio for superior boys is higher than that for superior girls.

3. The superior girls as a group were better adjusted socially and emotionally than were the superior boys.

4. Low achievement ratio is associated with intermediate, average, or normal personality adjustment for pupils of superior intelligence.

5. Poor personality adjustment is associated with highest achievement ratio, and best personality adjustment is associated with high average achievement ratio. The relation existing between personality and achievement is not positive, but tends to be negative. Pupils of superior intelligence who have well adjusted personalities have an advantage in academic competition over pupils with average personality adjustment, but pupils with poor personality adjustment have an advantage in achievement ratio over both the well adjusted and the average adjusted groups.

6. Of all the personality traits studied, truthfulness was the only one positively associated with achievement ratio. Truthful pupils were more likely to have high achievement ratios than untruthful pupils.

7. The relation to achievement ratio of such personality traits as happiness, alienation, sympathy, control, self-determination, economic self-determination, and particularly adjustment to the opposite sex all tended to be negative for superior pupils.

8. Superior pupils of high socio-economic status have an advantage in achievement ratio over pupils of low socio-economic status.

9. The lag between promise and performance of superior pupils of foreign extraction is less than that for pupils of native extraction.

10. Pupils in whose homes foreign languages are spoken have an advantage in achievement ratio over pupils in whose homes only English is spoken.

11. Pupils whose antecedents were born in Poland have in general an advantage in achievement ratio over pupils whose antecedents were born in Germany, Russia, or the United States.

12. In general, pupils whose parents and grandparents came from foreign countries consistently exceeded the mean achievement ratio of pupils whose parents and grandparents were born in the United States.

13. There is a positive association between achievement ratio and the amount of formal education of the pupil's father. The association exists, but it is not so well marked for the amount of formal education of the pupil's mother. The relation between achievement ratio and parents' education is greater for girls than it is for boys.

14. Superior pupils from small families of one or no siblings have an advantage in achievement ratio over pupils from large families.

15. Either the oldest or the youngest of the siblings or the only child has an advantage in achievement ratio over the intermediate siblings.

16. Girls from metropolitan areas have an advantage in achievement ratio over girls from rural areas.

17. Superior pupils, particularly boys, from broken families have an advantage in achievement ratio over pupils from intact families.

18. There is little relation between achievement ratio and pupil's health, number of diseases contracted, and physical defects. Cer-

tain throat and teeth defects had a slight influence on achievement ratio, reducing it a trifle.

19. Pupils whose parents' health was poor were likely to have higher achievement ratios than pupils whose parents' health was good.

20. Pupils who were deprived of privileges by their parents as a method of discipline had lower achievement ratios than pupils who were lectured, whipped, or punished in some other way. Pupils who were whipped or lectured had a higher mean achievement ratio than any other group.

21. Whether everybody or nobody in a family, broken or intact, worked has little effect on the achievement ratio of these superior pupils studied.

22. Good study habits are associated with a high achievement ratio for superior pupils, but the association is not marked.

A single thread appears to bind many of these conclusions together. Often, wherever a pupil of superior intelligence is found with a handicap, such as coming from a broken family, being of foreign extraction, poor health of parents, or poor personality adjustment, he works harder in school to overcome this handicap, and consequently his achievement ratio increases. In some cases, of course, this will not account for the existing relationship and other things may be responsible.

Implications for Further Research

There is available considerable additional information covering a large number of facts about this group of pupils with superior intelligence. The information covers sixty-six variable factors that may or may not be related to the achievement of superior pupils. At some future date, this same material may be used to compare the variables with one another or to compare all the variables with some variable besides achievement ratio.

A number of factors have been shown to be associated with the achievement ratio of superior pupils. Certain other variables have evidenced little or no relationship with the achievement ratio of superior pupils. On the basis of this study, it would seem worthwhile to investigate these relationships more thoroughly. The present research shows only that such a relationship is present. Further research might be helpful in determining the degree of and reasons for these significant relationships.

Other investigators may be interested in pursuing the investigation of the relationship to achievement of the conglomeration factor sometimes called "drive". There is evidently some third group of related variables besides intelligence and personality that has a direct effect on the achievement of not only superior pupils, but all pupils. Further research may tend to clear up just what this influence is.

Further research concerning how the school may better provide for the intellectually superior pupil could be based on such investigation as that done in this study. Applications of the truths concerning the factors associated with good achievement of superior pupils would be helpful. The school would not only like to know what factors are associated with the achievement of superior pupils, but would also like to know what program would encourage the development of those factors positively associated with achievement.

The pupil of superior intelligence deserves special attention. Generally and particularly in a democracy, the general level of learning of the population is no better tnan the level of the intellectual leaders. Tremendous efforts have been made to improve the achievement and learning level of the lower ability groups, so as to bring up the average level for the whole population. This is very desirable, but the same attention, if given to pupils of superior ability, would yield results ten-fold in value. The fertile soil of the superior child should not be neglected for the stony land of the low ability children. We are never any better than our leaders, and we should learn to train our best to be leaders.

Bibliography

Baker, Harry J. *Characteristic Differences in Bright and Dull Pupils.* Bloomington, Illinois: Public School Publishing Co., 1927. Pp. viii + 118.

"Baltimore Program for Pupils of Superior Ability," *Baltimore Bulletin of Education,* XVIII (September and October, 1940), 1–98.

Bell, Howard M. *Youth Tell Their Story.* A study conducted for the American Youth Commission. Washington: American Council of Education, 1938. Pp. v + 273.

"Characteristics of Superior Students," *National Education Association Research Bulletin,* XIX (September, 1941), 155–197.

Cohen, Helen Louise. "The Program for Gifted Pupils in New York City," *English Journal*, XXVI (September, 1937), 548–556.

Conklin, Agnes M. *Failures of Highly Intelligent Pupils: A Study of Their Behavior by Means of the Control Group*. Teachers College Contributions to Education, No. 792. New York: Teachers College, Columbia University, 1940. Pp. xii + 250.

Desing, Minerva F. *The Relation of Certain Pupil Achievement Gains to Certain Personal and Environmental Elements*. Philadelphia: University of Pennsylvania dissertation, 1940. Pp. xii + 169.

Eckert, Ruth E., and Mills, Henry C. "Variations among High School Seniors in Promise and Performance Measures," *School Review*, XLIV (April, 1936), 274–282.

Hollingsworth, Leta S. *Gifted Children: Their Nature and Nurture*. New York: Macmillan Co., 1926. Pp. xxiv + 374.

Jackson, Robert W. B. *Application of the Analysis of Variance and Covariance Method to Educational Problems*. Bulletin II. Toronto: Department of Educational Research of the University of Toronto, 1940. Pp. 103.

Jones, Harold E. "Relationships in Physical and Mental Development," *Review of Educational Research*, IX (February, 1939), 91–139.

Kelly, Truman L. "A New Method for Determining the Significance of Differences in Intelligence and Achievement Scores," *Journal of Educational Psychology*, XIV (September, 1923), 321–32.

Lawson, Edna E. "High School Achievement of Fifty-Six Gifted Children," *Pedagogical Seminary and Journal of Genetic Psychology*, XLVII (May, 1932), 326–30.

Lindquist, E. F. *Statistical Analysis in Educational Research*. Boston: Houghton Mifflin Co., 1940. Pp. viii + 266.

Link, Henry C., and Others. *Manual for the P.Q. or Personality Quotient Test: Inventory of Activities and Interests*. New York: Psychological Corporation, 1938 (revised). Pp. 14.

Myers, Charles E.; Ruch, Giles M.; and Loofbourow, Graham C. *Myers–Ruch High School Progress Test: Manual of Directions*. Yonkers-on-Hudson, New York: World Book Co., 1938. Pp. 12.

Newland, T. Ernest. "The Mentally Gifted," *Review of Educational Research*, XI (June, 1941), 277–287.

Purdom, T. Luther. "Traits Associated with the Success and Failure of Students," *The Non-Adjusted Pupil in Junior and Senior High School*, pp. 80–83. Second Yearbook of the Department of High School Principals. Lansing, Michigan: Michigan Education Assembly, 1931.

Regensburg, Jeannette. *Studies of Educational Success and Failure in Supernormal Children*. Archives of Psychology, No. 129. New York: Columbia University, 1931. Pp. 150.

Sims, V. M. *Score Card for Socio-Economic Status: Manual of Directions*. Bloomington, Illinois: Public School Publishing Co., 1927. Pp. 12.

Stagner, Ross. "The Relation of Personality to Academic Aptitude and Achievement," *Journal of Educational Research*, XXVI (May, 1933), 648–660.

Starrak, J. A. "Matching Ability With Achievement," *Journal of Higher Education*, VIII (June, 1937), 315–320.

Stroud, J. B. "Psychological Tests and Their Uses," *Review of Educational Research*, XI (February, 1941), 25–41.

Terman, Lewis M., and Others. *Genetic Studies of Genius: Mental and Physical Traits of a Thousand Gifted Children*. Stanford: Stanford University Press, 1925. Pp. xv + 648.

Traxler, A. E. "Current Construction and Evaluation of Personality and Character Tests," *Review of Educational Research*, XI (February, 1941), 57–79.

Washburne, John N. *Washburne Social Adjustment Inventory (Thaspic Edition): Manual for Interpreting*. Yonkers-on-Hudson, New York: World Book Co., 1940. Pp. 32.

Witty, Paul A. *A Study of One Hundred Gifted Children*. Lawrence, Kansas: Bureau of School Service and Research, University of Kansas, 1930. Pp. 44.

Wrenn, C. Gilbert, and McKeown, R. B. *Study Habits Inventory*. Stanford: Stanford University Press, 1934 (revised). Pp. 4.

AN EXPERIMENTAL STUDY OF TWO METHODS OF TEACHING CHEMISTRY IN SENIOR HIGH SCHOOL[1]

WILLIAM E. WARD
Fort Worth Public Schools

Until recently both teaching and teacher training have been aimed primarily at achieving mastery of subject matter. This objective has resulted in the development of a formal, mechanical classroom procedure that has emphasized the learning of subject matter for its own sake, with little regard for the final effect of this routinized procedure upon the learner himself. Today the curriculum revision movement is having a tremendous influence in bringing about a general change in point of view. We are now becoming more concerned with the welfare of the individual and of society. According to present-day educational philosophy, the growth of the individual must be fostered. Teaching methods should take into account individual differences and so vary the procedure as to fit the interests, needs, and abilities of the individual at various levels of maturation.

One result of the curriculum revision movement is the production and installation of numerous new course of study bulletins in various school systems throughout the nation. In general, these bulletins contain discussions of the present day educational philosophy based upon the idea of the complete growth of the whole child. They also contain materials intended to be helpful to the teacher in adapting subject matter and directing pupil experiences in such a way as to produce maximum growth of the individual, so that he may be an intelligent, cooperative member of the group, capable of meeting life situations effectively.

A question arises as to whether the methods resulting from the new courses are superior to the procedures previously used. In order that real progress may be made, it is important that the answer to this question be determined in each case, with scientific objectivity, through educational research.

[1] Abstract of a doctorate dissertation completed at the University of Texas in 1942, under the direction of Dr. James G. Umstattd.

THE PROBLEM

At the beginning of the school year of 1937–38, a new chemistry course of study[2] was installed in the Fort Worth high schools. In the following year, 1938–39, it was deemed advisable to test the value of the new course in chemistry. The writer, therefore, undertook the problem of comparing the new method of teaching chemistry, used in the high schools of Fort Worth as a result of the installation of the new course of study, with the traditional textbook method formerly used. It was believed that a comparison of this kind, conducted on a scientific basis, might serve to test the validity of the new course. This investigation sought to serve that purpose.

THE TWO METHODS

This investigation is an experimental study to determine the relative effectiveness of two methods of teaching chemistry at the secondary level. The two methods are referred to throughout the study as the traditional and the modern methods. Fundamentally the study might be thought of as being a comparison of two philosophies of teaching. Briefly stated, the traditional method is based upon the idea that subject matter is the end itself, while the modern method uses subject matter as a means to the end. The procedure resulting from the basic philosophy of the modern method consists of the selection of pupil activities approximating life situations and based upon the interests, needs, and capacities of the pupils. In this way it is believed the learning materials are made more meaningful and functional in the everyday lives of the pupils. The purpose of this study was to compare the results achieved through such an approach with those accomplished through the traditional textbook method.

A detailed discussion, outlining the specific procedures of the two methods, follows:

[2] Fort Worth Public Schools, *Science—A Tentative Course of Study for Chemistry*. Curriculum Bulletin No. 173, 1937.

The Traditional Method.—The traditional method, as agreed upon by the writer and the cooperating teachers in charge of the control groups, consisted of four distinct parts. These were:

1. The assignment of the lesson.
2. The recitation of the lesson.
3. The laboratory experiments.
4. Testing.

The entire procedure was carried on in a somewhat formal manner and the primary purpose was to have the pupil learn the subject matter. Neither the procedure for making lesson assignments nor the manner of conducting class recitations was standardized. Both of these procedures varied somewhat from day to day with the same teacher and also with different teachers. All of the teachers understood that they were to teach the traditional-method classes by the method they had used prior to the installation of the new course of study. The chief method of making assignments was to give the pupils a certain number of pages to prepare for the next lesson. Occasionally the teacher called attention to important aspects of the material to be studied and pointed out some difficulties with which the pupils might be confronted in the lesson. It was the general opinion of the cooperating teachers that when they taught in this way the classroom recitation procedure consisted mainly of the question and answer method and the lecture method. Both of these procedures were used primarily for the purpose of having the pupil learn the textbook material.

In the laboratory the procedure consisted chiefly of the individual method and the small-group method. In carrying out the experiments the pupils followed instructions prepared for them by the teacher. In some cases these instructions were written on the blackboard and in other cases they were mimeographed. Occasionally the laboratory demonstration method was used, either by the teacher or by a pupil appointed by the teacher to conduct the demonstration. In the laboratory the primary purpose of instruction was the learning of textbook information. After the experiments were completed the results were discussed. Pupils were questioned in order to bring the most important points to their attention. Teachers required pupils to keep notebooks in which they made records of the laboratory experiments and wrote balanced equations for the chemical reactions involved.

In the class recitation periods and also in the laboratory periods, practical applications of the subject matter were sometimes made. Limited reference was made to the use and importance of the scientific method. Very little was done with regard to developing an understanding and appreciation of the social significance of chemistry. Throughout the entire work, major emphasis was placed upon factual information. The first objective of the chemistry course was to have the pupils learn chemistry.

Regularly at the end of each third week the pupils were tested by means of a written test. This test was usually written on the blackboard and consisted primarily of questions concerning the subject matter of the textbook, chemistry problems, and the writing and balancing of chemical equations. Objective tests were sometimes given. The teacher prepared these tests to cover the subject matter studied. The three principal purposes of this testing procedure were to help the teacher discover the pupil's weaknesses in subject matter, to help the teacher find parts of the subject matter upon which he should place more emphasis, and to give the teacher an objective basis for marking the pupil.

The Modern Method.—In this method the teacher attempted to make use of pupil interests and needs to facilitate the learner's growth. By means of the pupil's school experiences with chemistry, the teacher should attempt to build up a comprehensive understanding of those major generalizations significant in everyday life. The major objective is that of achieving changed behavior on the part of the pupil. The direction this change should take is indicated by the objectives set up. The desired change in behavior can be brought about most effectively by providing experiences that will enable the pupil to understand science generalizations and to grow in ability to apply them to his daily life. The pupil should have experiences involving significant science concepts in lifelike situations with interesting content, and as a result of these experiences should achieve growth in the directions desired. The teacher is concerned primarily with "what" to teach and "how" to teach it. The former should be determined on the basis of the needs, inter-

ests, previous experiences, and abilities of the pupils. The latter should be determined on the basis of the best educational psychology. It follows then that subject matter is of necessity an integral part of the program, but should be considered as a means to an end. Also, each pupil should have opportunities to participate in purposing, planning, executing, and evaluating.

In 1938 a new suggestive course of study in chemistry[3] was issued by the Texas State Department of Education. In the modern method of the present study, the state course of study was used in connection with the Fort Worth course of study bulletin. Since the objectives set up in both of these bulletins are quite similar, it was decided to consider those listed in the state course as the accepted objectives of the modern method for the purposes of this study. They are:

To help the pupil:

1. To understand the nature of things and processes in daily use.
2. To learn how to avoid the danger of becoming victims of fraudulent advertisers and salesmen.
3. To acquire discrimination and habits of withholding judgment until the facts are known.
4. To become acquainted with scientific methods.
5. To understand the part chemistry plays in industrial and cultural progress (health, agriculture, and so forth).
6. To understand the relationship of science to other fields and to life in general.
7. To discover aptitudes and abilities in the field of chemistry.
8. To develop a foundation for college chemistry.[4]

In the new course of study[5] the teaching and learning materials are grouped under a unit organization. In the present study the units were developed on the basis of Thorndike's laws of learning and the basic philosophy of education. The cooperating teachers agreed upon a general classroom procedure, an outline of which follows:

[3] Texas State Department of Education, *Teaching Science*. Bulletin No. 383. Austin: State Department of Education, 1938.

[4] *Ibid.*, p. 36.

[5] Fort Worth Public Schools, *op. cit.*, p. 8.

Unit Development

I. Readiness

(a) Approach; (b) orientation; (c) planning.

The time necessary for this phase of the unit development will vary. It is suggested that from three to five days will ordinarily be needed. More time should be used if advisable for thorough work in order that the purposes be definitely achieved. This period is primarily concerned with creating a state of readiness, stimulating purposes, and planning the achievement of purposes.

1. Approach
   a. Stimulate desires to learn.
   b. Stimulate desires to achieve understanding.
   c. Insure permanent interest in the unit on the part of the pupil.
2. Orientation
   a. Acquaint the pupils with the unit as a whole.
   b. Bring the learners into contact with the general ideas involved.
   c. Determine the pupil experiential background.
   d. Discover definite needs in order that the pupils and the teacher may intelligently plan the activities.
3. Planning
   The pupils should participate in:
   a. What they want and need to know.
   b. What they will do in order to find out.
   c. How they will organize their efforts.

II. Exercise

The major part of the time consumed by a unit is usually used in this phase of the unit development.

1. Activities of this period.
   a. The pupils should modify and execute their plans.
   b. It is a period of work, drill, experiencing, and directed study.
   c. It should include: reading, listening, experimenting, demonstrating, arguing, criticizing, and so forth.
2. Everything attempted during this period should have a definite purpose that is focused upon achievement of desired understanding.

3. All materials and subject matter should contribute to and be focused upon the achievement of the understanding desired.

III. Effect[6]

This is a period of the unit development that has as its purpose the effective culminating of the unit.

It includes:

1. Assembling, organizing, and summarizing the activities.
2. Testing and self-testing. All of this testing need not be done at the end of the unit.
3. Providing corrective experiences, where needed.

It had been customary to think of the high-school science courses as being separated into biology, chemistry, and physics. In the modern method there is no intention of carefully drawing the lines of demarcation between these subjects. In actual classroom practice, perhaps the leading role in this procedure should very properly be played by integration. This integration must, of course, take place in the mind of the pupil, but the teacher can have a vital part in helping the process to go on. At every opportunity the teacher should attempt to help the child to integrate chemistry, not only with the other subjects, but also with life itself. This would result in pupils having experiences more nearly approximating life situations. Developing an understanding of the social significance of science should be useful in making the science materials more meaningful and functional, and therefore more significant in developing understandings desired.

Because of individual differences of both pupils and teachers, and also because of differences in local situations, the modern method places a great responsibility upon the teacher. The teacher must direct the procedure and adapt the method to the particular individual and group needs of the local situation. Therefore, actual practices of the teacher with regard to the use of this method will vary with the varying needs and interests of the pupils. In practice much individual instruction will be necessary. However, a greater amount of the instruction than one would ordinarily suppose may be accomplished in general class discussions and by dividing the class into smaller groups centered around common interests. While the modern method involves the pragmatic point of view, it aims at something higher than the mere utilitarian. Its purpose is to have every idea learned so that it will become a part of the living mind and aid in the interpretation of life. It recognizes the humanizing influence of freedom of creative expression, but also attaches major importance to guidance and direction. Its ultimate aim is the maximum development of the whole child.

[6] Because of the doubt in the minds of many, including Thorndike himself, regarding the accuracy of his interpretation of the law of effect, this law is here interpreted as meaning the actual application of the thing that results from the process of learning in life activity.

The Procedure

The study was conducted in twenty-one different chemistry classes in five high schools of the Fort Worth public school system. Five teachers and 420 pupils participated. The pupils did not know the study was being made. No names were used. Pupils were designated by Arabic numerals, teachers by capital letters, and classes by Roman numerals. The time devoted to the actual experimental work of the study was three semesters, namely, the second semester of the school year of 1938–39 and both semesters of the school year 1939–40. The entire work of the study, including time for preparation and making calculations and other details, was at least three full school years. The investigation naturally divided itself into three main parts, which may be indicated as:

I. The Preliminary Work

This division of the study began in the fall of 1938 and consisted of preparation and planning. By the end of the first semester, tests were prepared and mimeographed, the teachers had made the necessary preparation, and everything was in readiness for the actual experimental work to begin with the second semester in Chemistry II.

II. The Experimental Work

This division of the study consisted of the executing of the plans, the actual work of teaching and testing. This work was carried on for three semesters, namely, the second semester of the school year of 1938–39, and both the first and second semesters of the school year 1939–40. Thus, the experimental work was conducted in Chemistry I for one semester and in Chemistry II for two

semesters. Since the experimental work was carried on in three semesters with different groups, it may be divided into three phases:

1. First Phase.—In this phase of the experimental work the study was carried on in three high schools, included five classes and three teachers, and covered the work of the second semester in chemistry, Chemistry II. The primary purpose of this phase was the development of technique and procedure and the refining of tests.
2. Second Phase.—This phase of the study was conducted in five high schools, included nine classes and five teachers, and covered the work of the first semester in chemistry, Chemistry I.
3. Third Phase.—In this phase of the experimental work four high schools were involved, including seven classes and four teachers, and covering the second semester of chemistry, Chemistry II.

III. The Mathematical Treatment

This division of the study consisted of collecting and assembling data, checking results, calculating coefficients of correlation, validating tests, calculating and comparing means, medians, standard errors, and critical ratios, and drawing conclusions.

The equivalent-group method was used. Classes were equated on the basis of intelligence quotient, scholarship average, and scores made on two tests. One of these tests was a non-informational or functional test, and the other was a test of chemistry information. The functional test attempted to measure progress in development of a scientific attitude, practical application of chemistry to everyday life, and an understanding of the social significance of chemistry. Twenty pupils in each class were paired with twenty pupils in the other class by the matching process. One of the classes of each pair of equated classes taught by the same teacher was instructed by the traditional method and the other by the modern method. At the end of the semester the same two tests were given again and the gains on scores made on the second testing over those made on the initial testing were calculated. These gains were ranked and the mean and median found for each class. The means and medians were then compared for each pair of equated classes. This was done by calculating the standard error for each measure and the standard error of the difference of each pair of measures. The method outlined by Gray and Votaw[7] was used. The critical ratios were then computed by dividing the difference between each pair of measures by the standard error of the difference. If the quotient thus found was three or more, the difference was considered as being significant and not a mere chance difference.

In this way the pairs of equated classes taught by the same teacher, using one method in the first class and the other method in the second class, were compared. In another set of comparisons one class taught by the modern method was used as a standard and equated with each of the other classes taught by other teachers. In this manner the modern method as taught by one teacher was compared with the traditional method as taught by other teachers. Also the modern method as taught by one teacher was compared with all other modern-method classes as taught by other teachers. Another type of comparison was made by equating a traditional-method class as a standard with all other traditional-method classes taught by other teachers. If significant differences are found to exist when modern-method classes are compared with traditional-method classes, but no significant differences are found either when modern-method classes are compared with other modern-method classes or when traditional-method classes are compared with other traditional-method classes, this would be very strong evidence of a real difference between the two methods.

The comparative value of the two methods as they relate to intelligence was found by comparing the scores made by the highest fifteen per cent of 160 pupils, ranked on the basis of intelligence quotients and taught by one method, with the highest fifteen per cent of 160 pupils ranked in the same way and taught by the other method. This was also done for the lowest fifteen per cent.

The Tests

In developing the tests for this study, the objectives of the course were kept constantly in mind and the tests were constructed with the purpose of measuring progress made in the direction of attainment of these objec-

[7] C. T. Gray and D. F. Votaw, *Statistics Applied to Education and Psychology*, pp. 191–192. New York: Ronald Press Co., 1939.

tives. With one exception, all tests used in this study were constructed by the investigator. This one exception was the Otis Group Intelligence Test, which is a standard test of well-known validity and reliability. It was used, therefore, as a criterion in determining the efficiency of the investigator's tests. As previously stated, two types of tests were used: the chemistry information tests for Chemistry I and for Chemistry II, and the non-informational or functional test. These tests were given to 140 pupils and when marked the reliability was found by the split-half technique. All of the tests used in the study were found to have a reliability coefficient above point nine. Information concerning the validity of the tests was found by calculating the correlation between scores on the tests and various criteria. These criteria were: scores on intelligence test, scholarship average, grade average made on previous science courses, and weights given for number of years of previous science. The product-moment method was used for calculating these coefficients of correlation.

Findings

The critical ratios were calculated for differences between the pairs of means and of medians, for each pair of equated classes compared in all three phases of the study. The results of these calculations are presented in Tables I, II, III, and IV. These tables show the classes compared and the D/σ_D values on both types of tests; namely, chemistry informational and the non-informational or functional tests. The tables of critical ratios are divided into three main columns, one for each of the three phrases of the study. Each of these main columns has two divisions in which the critical ratios of means and of medians are shown for both the functional and chemistry informational types of tests.

In Table I are found the pairs of equated classes taught by the same teacher, using the traditional method with one class and the modern method with the other class. In this group there are nine such pairs of classes in the entire study: two in the first phase, four in the second phase, and three in the third

TABLE I

RESULTS OF COMPARISONS OF TRADITIONAL-METHOD CLASSES WITH MODERN-METHOD CLASSES TAUGHT BY THE SAME TEACHERS

Class	Method	Teacher	First Phase Chem. II		Second Phase Chem. I		Third Phase Chem. II	
			Test III Functional	Test IV Chem. Inform.	Test II Functional	Test I Chem. Inform.	Test II Functional	Test III Chem. Inform.
			Critical Ratios for Medians					
II	T	B	.32	—0.25	3.1	3.1	1.7	1.9
III	M	B						
IV	T	C	1.00	—2.46	.8	2.7	1.7	1.6
V	M	C						
VI	T	D						
VII	M	D	------	------	3.9	4.8	2.1	------
VIII	T	E						
IX	M	E	------	------	.1	5.3	------	------
			Critical Ratio for Means					
II	T	B						
III	M	B	1.97	—0.20	4.9	2.7	1.7	1.7
IV	T	C						
V	M	C	2.28	—2.79	3.0	4.9	1.2	2.7
VI	T	D						
VII	M	D	------	------	4.8	5.1	1.5	—1.0
VIII	T	E						
IX	M	E	------	------	.9	5.4	------	------

Note: "T" represents Traditional Method.
"M" represents Modern Method.
(—) quantities favor the Traditional Method.
Critical Ratio = D/σ_D.

TABLE II

RESULTS OF COMPARISONS OF TRADITIONAL-METHOD CLASSES WITH MODERN-METHOD CLASSES TAUGHT BY DIFFERENT TEACHERS

Class	Method	Teacher	Critical Ratios for Medians					
			First Phase Chem. II		Second Phase Chem. I		Third Phase Chem. II	
			Test III Functional	Test IV Chem. Inform.	Test II Functional	Test I Chem. Inform.	Test II Functional	Test III Chem. Inform.
I	M	A						
II	T	B	1.79	3.75	3.8	2.1	3.9	6.1
I	M	A						
IV	T	C	3.07	3.02	2.1	2.4	5.0	2.8
I	M	A						
VI	T	D	------	------	4.5	3.6	3.5	1.7
I	M	A						
VIII	T	E	------	------	—3.1	3.0	------	------
			Critical Ratios for Means					
I	M	A						
II	T	B	3.11	4.08	4.4	4.3	3.1	5.4
I	M	A						
IV	T	C	4.68	2.87	4.3	3.4	4.9	4.9
I	M	A						
VI	T	D	------	------	5.3	6.7	3.9	1.1
I	M	A						
VIII	T	E	------	------	—2.3	4.8	------	------

Note: "T" represents Traditional Method.
"M" represents Modern Method.
(—) quantities favor the Traditional Method.
Critical Ratio = D/σ_D.

phase. An inspection of Table I shows that there are thirty-six D/σ_D values obtained for both mean and median scores on the two types of tests, eighteen for the means and eighteen for the medians. Of the thirty-six critical ratios, eleven reveal significant differences in favor of the modern method. Six of these eleven are differences of the means and five are differences of the medians. All of these significant differences are found in the second phase of the study. There are no significant differences in favor of the traditional method. The remaining twenty-five of the thirty-six values are too small to indicate a true difference.

In Table II are found the pairs of equated classes taught by different teachers, using the traditional method with one class and the modern method with the other class. In this group there are nine such pairs of classes in the entire study: two in the first phase, four in the second phase, and three in the third phase. An inspection of Table II shows that there are thirty-six D/σ_D values obtained for both mean and median scores on the two types of tests, eighteen for the means and eighteen for the medians. Of the thirty-six critical ratios, twenty-seven reveal significant differences. Fifteen of these twenty-seven are differences of the means and twelve are differences of the medians. The remaining nine differences are too small to be significant. There is one significant difference in favor of the traditional method. This one case is found in the second phase of the study as a result of comparing the medians made by classes I and VIII on the functional test.

Table III includes the cases where modern-method classes are compared with other modern-method classes. These pairs of equated classes are taught by different teachers and in this division will be found nine such pairs of equated classes for the entire study. Two of these nine are in the first phase, four in the second phase, and three in the third phase of the study. An inspection of Table III reveals that there are thirty-six D/σ_D values obtained for both mean and median scores on the two types of tests, eighteen for the means and eighteen for the medians. Of the thirty-six critical ratios, ten reveal significant differences. The remaining

TABLE III

RESULTS OF COMPARISONS OF MODERN-METHOD CLASSES WITH OTHER MODERN-METHOD CLASSES TAUGHT BY DIFFERENT TEACHERS

Class	Method	Teacher	Critical Ratios for Medians					
			First Phase Chem. II		Second Phase Chem. I		Third Phase Chem. II	
			Test III Functional	Test IV Chem. Inform.	Test II Functional	Test I Chem. Inform.	Test II Functional	Test III Chem. Inform.
I	M	A						
III	M	B	2.11	**4.56**	.8	.6	1.0	**4.4**
I	M	A						
V	M	C	**3.00**	**4.95**	2.7	.3	**5.2**	1.4
I	M	A						
VII	M	D	------	------	.4	.1	2.1	2.0
I	M	A						
IX	M	E	------	------	.8	2.6	------	------
			Critical Ratios for Means					
I	M	A						
III	M	B	1.64	**5.01**	.7	1.1	1.5	**5.1**
I	M	A						
V	M	C	**3.77**	**4.38**	1.9	.4	**5.5**	.9
I	M	A						
VII	M	D	------	------	.8	2.2	1.5	1.8
I	M	A						
IX	M	E	------	------	1.2	2.9	------	------

Note: "T" represents Traditional Method.
"M" represents Modern Method.
(—) quantities favor the Traditional Method.
Critical Ratio = D/σ_D.

TABLE IV

RESULTS OF COMPARISONS OF TRADITIONAL-METHOD CLASSES WITH OTHER TRADITIONAL-METHOD CLASSES TAUGHT BY DIFFERENT TEACHERS

Class	Method	Teacher	Critical Ratios for Medians					
			First Phase Chem. II		Second Phase Chem. I		Third Phase Chem. II	
			Test III Functional	Test IV Chem. Inform.	Test II Functional	Test I Chem. Inform.	Test II Functional	Test III Chem. Inform.
II	T	B						
IV	T	C	1.4	1.3	1.3	.7	.15	**3.4**
II	T	B						
VI	T	D	------	------	**2.1**	1.7	1.2	**3.08**
II	T	B						
VIII	T	E	------	------	**6.0**	.1	------	------
			Critical Ratios for Means					
II	T	B						
IV	T	C	1.1	.5	.2	.09	1.1	2.2
II	T	B						
VI	T	D	------	------	1.9	1.4	.9	**3.0**
II	T	B						
VIII	T	E	------	------	**6.3**	2.4	------	------

Note: "T" represents Traditional Method.
"M" represents Modern Method.
(—) quantities favor the Traditional Method.
Critical Ratio = D/σ_D.

twenty-six differences are too small to be significant. None of these significant differences is found in the second phase of the study.

In Table IV are the results of comparing traditional-method classes with other traditional-method classes as taught by different teachers. In this category six pairs of equated classes are used for the entire study: one in the first phase, three in the second phase, and two in the third phase. An inspection of Table IV discloses twenty-four D/σ_D values obtained for both mean and median scores on the two types of tests, twelve for the means and twelve for the medians. Of the twenty-four critical ratios, five reveal significant differences. Two of the five cases are found in the second phase of the study, for the means and medians made on the functional test by classes II and VIII.

In considering the results of the comparison of the modern method and the traditional method in relation to intelligence, an inspection of Tables V, VI, and VII reveals the following: The modern method gave superior results with the lowest I. Q. group as measured by both Tests I and II. For the results of Test I on chemistry information over Chemistry I, two of the three comparisons for the lower I. Q. group yielded significant differences in favor of the modern method, as shown in Table V. The results from the functional test, shown in Table VI, reveal a similar result, two of the three critical ratios being greater than three. On Test III the differences were not significant. The modern method and traditional method seem to be about equal with respect to the highest I. Q. groups, as measured by all three tests, since none of the critical ratios indicates a significant difference.

With regard to discipline, all teachers report fewer disciplinary cases in their modern-method classes than in the traditional classes. The writer does not wish to give the impression that the modern method is perfect or ideal. Disciplinary problems did arise in the modern-method classes. A great deal depends upon the resourcefulness of the teacher and his ability to use the modern method. As long as pupils are interested in doing something for which they experience a need, there is no question of discipline.

Conclusions and Implications

When drawing conclusions upon the basis of the results of the experimental work, it should be borne in mind that the first phase of this study was a preliminary part, with the primary purpose of developing procedures and techniques for conducting the second and third phases of the study more scientifically. It seems reasonable to believe that the second phase of the study is likely to give the most accurate results, for in this phase the experimental study was carried on under the best and most favorable conditions. All pupils involved in this phase were beginning Chemistry I and none of them had studied chem-

Table V

Comparison of Mean Score and Mean Gain on Test I Made Under the Two Methods by the Highest and the Lowest Fifteen Per Cent of 160 Pupils, When Ranked According to Intelligence Quotients

Groups Compared	Measures Compared	Modern Method	Traditional Method	Critical Ratios
Upper Fifteen Per Cent	Mean I. Q.	114.0	117.0	.5
	Mean Score	77.5	71.0	1.9
	Mean Gain	63.0	51.0	2.6
Lower Fifteen Per Cent	Mean I. Q.	94.0	95.0	.1
	Mean Score	56.0	31.0	4.2
	Mean Gain	47.0	21.0	5.7

Note: "Score" refers to final score made on Test 1. "Gain" refers to difference between initial score and final score made on Test I.

TABLE VI

COMPARISON OF MEAN SCORE AND MEAN GAIN ON TEST II MADE UNDER THE TWO METHODS BY THE HIGHEST AND THE LOWEST FIFTEEN PER CENT OF 160 PUPILS, WHEN RANKED ACCORDING TO INTELLIGENCE QUOTIENTS

Groups Compared	Measures Compared	Modern Method	Traditional Method	Critical Ratios
Upper Fifteen Per Cent	Mean I. Q.	112.0	115.0	.4
	Mean Score	79.4	79.0	.0
	Mean Gain	9.0	10.5	.3
Lower Fifteen Per Cent	Mean I. Q.	94.0	95.0	.1
	Mean Score	77.0	68.0	3.2
	Mean Gain	11.0	5.0	3.9

Note: "Score" refers to final score made on Test II. "Gain" refers to difference between initial and final scores made on Test II.

TABLE VII

COMPARISON OF MEAN SCORE AND MEAN GAIN ON TEST III MADE UNDER THE TWO METHODS BY THE HIGHEST AND THE LOWEST FIFTEEN PER CENT OF 140 PUPILS, WHEN RANKED ACCORDING TO INTELLIGENCE QUOTIENTS

Groups Compared	Measures Compared	Modern Method	Traditional Method	Critical Ratios
Upper Fifteen Per Cent	Mean I. Q.	112.0	114.0	.3
	Mean Score	38.5	39.3	.1
	Mean Gain	21.7	23.6	.3
Lower Fifteen Per Cent	Mean I. Q.	94.0	95.0	.5
	Mean Score	29.0	26.0	1.4
	Mean Gain	15.1	14.6	.2

Note: "Score" refers to final score made on Test III. "Gain" refers to difference between initial and final scores made on Test III.

istry previously. The plans for carrying on the study had been developed, as had the technique to be used. This was also true of the third phase, but in the third phase of the investigation the personnel of each class had been changed to a great extent. This was due to changing individual schedules and programs at midyear. In the third phase of the study other conditions arose that affected the accuracy of the results.

When only the second phase is considered, the results are almost completely in favor of the modern method. Where modern-method classes are compared with other modern-method classes, and traditional-method classes are compared with other traditional-method classes, the results show them to be practically equal. Yet, when modern-method classes are compared with traditional-method classes, there are numerous cases of significant differences favoring the modern method.

Where modern-method classes are compared with traditional-method classes taught by the same teacher, all differences on both functional and non-functional tests favor the modern method and eleven of the sixteen dif-

ferences in this category are significant. Where modern-method classes are compared with traditional-method classes taught by different teachers, there are sixteen cases in all, including means and medians on both the functional and non-functional tests. Of these sixteen cases, eleven show significant differences favoring the modern method and there is one significant difference in favor of the traditional method. This is the comparison of Classes I and VIII on medians for the functional test.

When modern-method classes are compared with other modern-method classes, no significant differences are found and when traditional-method classes are compared with other traditional-method classes two of the twelve comparisons in this category are significant differences; these are in favor of Class VIII for both means and medians on the functional test. Teacher "E" accounts for this difference by saying that he frequently found himself emphasizing the social significance of chemistry in Class VIII, which was a traditional-method class. Owing to his enthusiasm and to the habit of teaching in this manner, he often forgot to adhere strictly to the traditional method in this class. This is no doubt the real explanation, for this difference holds true with regard to Class VIII in every case in which it is compared with other classes. For example, when Classes VIII and IX are compared on the functional test, very slight differences are found for both means and medians. These classes were taught by teacher "E", using different methods. In the second division, when Class VIII, a traditional method class taught by teacher "E", is compared with Class I, a modern method class, taught by teacher "A", a significant difference for the medians favoring the traditional method is found, and a high difference (—2.3) of the means, though not significant, is found to favor the traditional method. This, in itself, would not prove the point, if it were not for the fact that all other values in this group are in favor of the modern method, and most of them are significant. This is a good illustration of teacher differences.

All of the following conclusions, drawn on the basis of evidence found in this investigation, refer to the comparison of the methods, as taught by these teachers and as measured by these tests.

1. The modern method seems to be superior to the traditional method for general class work with heterogeneous groups.

2. Selection of learning materials on the basis of interests, needs, and capacities is superior to the use of subject matter on the traditional basis.

3. With regard to intelligence, evidence obtained in this study seems to show that the modern method is equal to the traditional method for pupils of high I. Q.'s and is superior to the traditional method for pupils of lower levels of intelligence.

4. Pupils of high I. Q.'s learn equally well by either method, as measured by both the chemistry informational and the functional tests.

5. Pupils of lower I. Q.'s learn better by the modern method than by the traditional method, as measured by these tests.

6. Pupils learn not only the facts and principles of chemistry better by the modern method, but also learn to appreciate the social significance of chemistry better when taught by this method.

7. In general, more transfer is obtained by the modern method, as indicated by the application of chemistry to practical life situations.

8. In general, the scientific attitude is developed better by the modern method than by the traditional method.

9. The modern method is superior to the traditional method as these methods relate to disciplinary problems. Where pupils are interested in accomplishing something they wish to do, are able to achieve, and feel a need for doing, disciplinary problems vanish. The degree to which they disappear depends upon the teacher's ability to use the modern method.

Significant implications that follow from these conclusions are:

1. Texts should be used as means to an end rather than as the end itself.

2. Teachers should think more of what the subject matter does to the child and perhaps less of what the child does to the subject matter.

3. Teachers must begin with the child where he is and take him as far as they can along the road to maximum growth and development.

4. In addition to trying to stimulate interest, the science teacher should try to locate already present interests and relate these to new situations, so as to broaden the child's outlook and understanding.

5. The teacher should use numerous illustrations and analogies, so as to relate the learning materials to a great variety of life situations. In this way integrative teaching may result in transfer of training. Emphasis should be placed upon basic principles, concepts, and generalizations.

6. Teachers of chemistry should discontinue the use of the traditional method as such, and should try to develop a method and technique patterned, at least in part, after the modern method based upon the interests, needs, and capacities of the pupils. It is evident that this should help to improve instruction and aid in making more and better progress toward our goal; namely, the maximum growth and development of the whole child for the best interests of both the individual and society.

A problem suggested to the investigator by the present study concerns the measuring and comparing of achievement resulting from different methods of instruction. In this study the gains in final scores over initial scores were compared. It seems to the writer that this method is more accurate than simply comparing the final scores. When only the final scores are compared, an error is involved, owing to previous growth not achieved by the methods of instruction being compared. However, when gains are compared, a question arises relating to differences in proportion of possible achievement that may be measured by the tests used. The writer is planning further investigation into this aspect of the present study.

THE RELIABILITY AND PERMANENCE OF VOCATIONAL INTERESTS OF ADOLESCENTS

KATHERINE VAN FRIDAGH TAYLOR[1]
University of California
Institute of Child Welfare

INTRODUCTION

In recent years considerable emphasis has been placed in the secondary school upon vocational guidance to the end that more adequate vocational adjustments might subsequently be effected. As is not uncommon in educational programs, the theory and objectives have far outstripped available techniques for achieving stated ends.

Although consistently questioned by workers in the field, vocational choices have usually provided the basis for guidance insofar as the interests of the individual are concerned. The continued use of such methods has been due not so much to the fact that their limitations have not been recognized as to the absence of an objective and reliable tool for measuring vocational interests.

Since 1920, many investigators in this field have been concerned with the construction and perfection of devices for the measurement of interests, employing objectively scored inventories standardized on adult persons employed in various occupations. The norms that have been secured in this fashion are, consequently, not intended for use with high school pupils. The present study was undertaken to determine the suitability of such a technique at the senior high school level. The motive behind the investigation has been a belief in the necessity for adequate counseling data, available to the individual at such an age that the impact of vocational failure or unadjustment may be avoided and time and money may be husbanded in securing professional and vocational preparation.

The adaptation to adolescents of an inventory standardized for measuring vocational interests of adults depends chiefly upon three main characteristics of adolescent interests. In the first place, interests susceptible to such measurement must be present to a sufficient degree to permit diagnosis. Secondly, the instrument used to measure such interests must be reliable at the age level under consideration. Thirdly, the interests measured must exhibit a reasonable degree of permanence and stability, if the results of such measurement are to be used by the counselor for the direction of vocational and educational programs.

[1] The writer is indebted to Dr. Mary Cover Jones, who assisted in the collection of the data, and to Dr. Herbert S. Conrad and Dr. Harold D. Carter, who read and criticized the manuscript.

THE DATA AND SAMPLE

The data were secured by administration of the Strong Vocational Interest Blanks for Men and Women to sixty-four high school boys and sixty-two high school girls, respectively. The testing period covered four years, from an average age of sixteen years (15.9 years for boys and 15.7 years for girls) at the first testing to twenty years (19.9 years for boys and 19.8 years for girls) at the time of the fourth test. The girls were tested first with the Men's Blank and then with the Women's Blank on the three subsequent tests. Ten vocational scales selected as roughly representative of the major occupational fields available were scored for each group at each testing.

The sample was selected as representative of an urban senior high school population.

During the first three years of the testing program, the group remained in high school. At the time of the fourth test they had been graduated from high school for a year and a half. Sixty-three per cent of the boys were continuing in school, thirty-four per cent were employed, and three per cent were looking for work. Of the girls, thirty-seven per cent were in school, thirty-six per cent were employed, sixteen per cent were married and keeping house, and eleven per cent were at home with their parents.

METHOD OF ANALYSIS AND RESULTS

The first step in the analysis of the data was the standardization of all scores in order to improve ease of handling and to secure

comparability of scores from scale to scale. When this had been accomplished, the seventeen-year test (the second test) for boys was rescored in odd and even halves, and coefficients of reliability for the ten scales were computed. These coefficients are presented with those obtained by Strong on an adult sample through the use of the same method (Table I).

TABLE I

COMPARISON OF RELIABILITY COEFFICIENTS OBTAINED FROM SCORES OF HIGH SCHOOL BOYS WITH THOSE OBTAINED FROM ADULT SCORES

Scale	High School Boys Mean Age— 16.7 yrs.	Adults[1]
Physicist	.90	.94
Chemist	.90	.89
Physician	.88	.88
Lawyer	.86	.90
Life Insurance Salesman	.88	.90
Real Estate Salesman	.75	.90
Teacher	.82	.90
Y. M. C. A. Secretary	.95	.88
Office Worker	.89	.89
C. P. A.	.81	.74

[1] Taken from E. K. Strong, Jr., "Permanence of Vocational Interests," *Journal of Educational Psychology*, XXV (1934), 336–44.

The same treatment of the data obtained from the second test administered to the girls yielded similar results (Table II).

TABLE II

COMPARISON OF RELIABILITY COEFFICIENTS OBTAINED FOR ADOLESCENT GIRLS WITH THOSE OBTAINED FOR ADULT WOMEN

Scale	High School Girls Mean Age = 17 yrs. N = 62	Married Women[1] Mean Age = 38 yrs. N = 500
Librarian	.95	.87
Physician	.91	.87
Nurse	.88	.87
Social Worker	.94	.83
Teacher	.88	.86
Artist	.96	.93
Lawyer	.82	.81
Life Insurance Saleswoman	.78	.74
Office Worker	.91	.92
Stenographer-Secretary	.92	.85

[1] Taken from E. K. Strong, Jr., *Manual for Vocational Interest Blank for Women*, p. 16. Stanford, California: Stanford University Press, 1938.

The obtained reliabilities are fairly adequate for individual prediction except, in the case of the boys, for the Real Estate Salesman scale and, in the case of the girls, for the Life Insurance Saleswoman scale. It is apparent, moreover, that the reliability coefficients obtained for adolescents compare favorably in most instances with those obtained for older persons. It will be observed that the split-half technique yielded somewhat higher reliabilities for the girls than those obtained for the boys. Whether this result is a function of the test, a true sex difference at this age level, or the result of lower differential weights on the scoring keys has not been investigated.

Attention was then turned to the question of the magnitude of the obtained scores, that is, the presence of measurable interests to a degree sufficient for guidance. The technique employed for this purpose was the analysis of letter ratings received by adolescent subjects in comparison with those received by the adult samples upon whose scores the tests were standardized (Table III).

TABLE III

DISTRIBUTION OF LETTER RATINGS FOR ADULT MEN AND HIGH SCHOOL SOPHOMORE BOYS AND GIRLS, EXPRESSED IN PERCENTAGES

Rating	Adult Men (After Strong)[1]	High School Boys	High School Girls
A	4%	6%	22.9%
B	29%	24%	23.4%
C	67%	70%	53.7%

[1] Taken from E. K. Strong, Jr., *Manual for Vocational Interest Blank for Men*. Stanford, California: Stanford University Press, 1937.

This comparison indicates the presence of measurable interests to an extent equal to those shown by adults. In a comparison of the data for boys and girls, the most noticeable discrepancy is found in the percentages of ratings received, 22.9% "A's" for the girls contrasted with 6% "A's" for the boys. Inspection of Table IV will indicate the probable cause for this difference.

Such a situation as that found on the Nurse, Artist, Office Worker, and Stenographer-Secretary Scales does not exist for the boys. The Housewife Scale is excluded from consideration on this point, since the obtained result needs no explanation. The high percentage of "A" and "B+" ratings received

TABLE IV

COMPARISON OF PER CENT OF "A" AND "B+" RATINGS ON STRONG VOCATIONAL INTEREST BLANK FOR WOMEN OF PROFESSIONAL SAMPLING, CONTROL GROUP, AND HIGH SCHOOL GIRLS

Scale	Professional [1]		Control [1]		High School	
	No.	%	No.	%	No.	%
Librarian	427	84.3	500	6.6	73	6.9
Physician	195	82.6	500	3.8	73	2.8
Nurse	396	83.9	500	45.4	73	54.8
Social Worker	177	84.8	500	6.4	73	2.8
Social Science Teacher	194	86.1	500	12.2	73	0.0
Artist	184	84.3	500	9.8	73	26.0
Lawyer	159	83.6	500	5.8	73	4.1
Life Insurance Saleswoman	77	81.8	500	1.6	73	1.4
Office Worker	226	83.2	500	67.4	73	78.2
Stenographer-Secretary	298	82.6	500	54.6	73	83.6
Housewife	500	84.2	None	------	73	84.9

[1] Taken from E. K. Strong, Jr., *Manual for Vocational Interest Blank for Women.* Stanford, California: Stanford University Press, 1938.

is not a function of the sample, since comparable ratings were secured by Strong's five hundred married women who constitute the control group. Whether an important sex difference exists here in the expression of several general and fundamental interests is obscured by the primary assumption which must be made, namely, that the discrimination of the women's scoring keys breaks down seriously on these four scales.

It will be noticed that the seventeen-year-old girls received high ratings in as great a proportion as did the adult women, except on scales for Social Worker and Social Science Teacher. Comparisons may not be drawn with the professional groups because of the difference in method of selecting the sample.

The presence and reliability of interests having been established, the value of the Strong Blanks is seen to depend upon the permanence of obtained interest scores. Several techniques for measuring stability have been used. Of these, test-retest correlations with length of interval and age as variable factors were emphasized, since the continuous sample provided opportunity for exploration of the developmental aspects of permanence coefficients.

It was found that increasing intervals progressively lowered test-retest coefficients for both boys and girls. When length of interval was held constant for boys, it was seen that stability regularly increased with age (Table V).

It is apparent from Table V that the age factor markedly influences the stability of the test-retest scores. To say "age factor" is misleading, perhaps, particularly since no environmental factors have been taken into consideration. The higher stability coefficients for Tests 3 and 4 may be due to age, maturity, experience, training; it is impos-

TABLE V

STABILITY OF SCALE SCORES AT INCREASING AGE LEVELS WITH LENGTH OF INTERVAL CONSTANT: N = 64

Scale	Test 1 & 2 1 yr. int. Age 16–17	Test 2 & 3 1 yr. int. Age 17–18	Test 1 & 3 2 yr. int. Age 16–18	Test 3 & 4 2 yr. int. Age 18–20
Teacher	.57	.70	.49	.73
Y. M. C. A. Secretary	.60	.80	.55	.72
Office Worker	.60	.68	.48	.62
C. P. A.	.64	.72	.57	.63
Real Estate Salesman	.53	.56	.52	.59
Physicist	76	.79	.70	.82
Doctor	.78	.83	.61	.79
Chemist	.71	.66	.65	.80
Lawyer	.71	.69	.59	.72
Life Insurance Salesman	.69	.67	.66	.77
Masculinity—Femininity	.73	.71	.73	.75
Studiousness	.55	.66	.54	.66
Interest Maturity	.71	.76	.67	.82

sible to say specifically which one or which combination of factors may be responsible, since the only apparent difference is one of age. It is known, concerning this sample, that during the time of the first three tests the boys were in high school. The interval between Tests 3 and 4 has been filled by semi-skilled jobs or by general academic and business training. Since this is so, if specific environmental factors such as training and experience are expected to have a significant effect upon stability, one should have anticipated a drop in the permanence coefficients for Tests 3 and 4, due to the marked change in activities that took place in the interval between these two tests. If, however, stability of interest response is due to factors in the nature of emergent personality traits and attitudes, one may accept the increase in stability coefficients as a development largely dependent upon age or maturity. The size of the sample does not permit any broad generalization regarding permanence of interests or developmental trends inherent in the stabilization process.

Nevertheless, the obtained coefficients compare favorably with those secured by other investigators who used subjects who were, on the average, three years older (7, 13). Coefficients were not, however, as high as those secured upon adult samples averaging six or more years older who were tested as seniors in college and again five years later (9).

The results for the girls are not as clear-cut, nor are the data as adequate. Since the testing period was shorter than that for boys, no comparison at different age levels with the interval held constant is possible. However, although the test instruments are different, making direct comparison impossible, the averages of the initial reliabilities of the two tests, .88 and .87, are close enough to permit argument by analogy. To do this, age level and interval must be held constant. When this is done, it appears that on tests administered at 17 years and 18 years, a one-year interval, the girls achieved an average stability coefficient of .74, and the boys, .72. On tests given at 18 years and 20 years, a two-year interval, the girls' average dropped to .66, while the boys' average increased to .73. There is no development of permanence of interests in the case of girls such as that evidenced by the boys. Why this should be so is not clear, unless it may be attributed to the fact that 27 per cent of the girls were married or at home preparing for marriage, whereas the boys were looking forward to a vocational career of some sort. This explanation is borne out by the increasing stability of Femininity-Masculinity responses, which has been accompanied by an increasingly higher mean score on this scale at each testing. If such a conclusion is acceptable, it would tend to bear out the popular contention that the interests of women are less channelized and not as intense as those of men.

Analysis of letter ratings as a method of measuring stability was also employed, since it is by means of such ratings that the inventory scores are generally interpreted. An analysis of letter ratings might, therefore, reveal results more nearly comparable with available evidence of adult reactions.

When the analysis of ratings was completed, such a result was obtained. Expressed in terms of the percentage possibility of obtaining a given score on a retest at a stated interval, the probabilities were found to be slightly greater for both boys and girls than for the adult samples with which comparisons were made (7, 13). Two of the studies reported were based on the administration of the Strong Vocational Interest Inventory to a sampling of college students with an interval of three years between test and retest. The third study was based on an interval of five years between tests (7).

Table VI indicates that 69.4 per cent of the ratings received by the boys in 1937 were duplicated after the four-year interval. Table VII provides similar data for the girls; 65.9 per cent of the ratings obtained by the girls in 1938 were identical with those received after an interval of three years. Strong (7), in reporting the results of a five-year study of college graduates, stated that 63.3 per cent of the ratings remained the same. Of course, the longer interval employed by Strong must be considered in any comparison.

Vocational interests, as measured by the Strong inventories, appear to be almost as permanent during high school years as in adult life. Whether the stability is great enough to warrant the use of the instrument for guidance is another question. Such use of these interest inventories seems not unrea-

TABLE VI

DISTRIBUTION OF 10 OCCUPATIONAL INTEREST RATINGS RECEIVED BY BOYS IN 1937 AND 1941, EXPRESSED IN PERCENTAGES: N = 64

Occupational Interest Rating	% obtained	% of those receiving a given rating in 1937 who received the same or another rating in 1941				
		C	B—	B	B+	A
A	6.02%	.33%	0.0%	.50%	1.34%	3.85%
B+	6.86	.33	.33	1.85	2.68	1.67
B	12.54	4.01	2.34	3.68	1.51	1.00
B—	4.85	2.51	.67	1.17	.33	.17
C	69.73	58.52	4.18	4.18	1.85	1.00
Ratings, 1941	100.00%	65.70	7.52	11.38	7.71	7.69%

TABLE VII

DISTRIBUTION OF 11 OCCUPATIONAL INTEREST RATINGS RECEIVED BY GIRLS IN 1938 AND 1941, EXPRESSED IN PERCENTAGES: N = 62

Occupational Interest Rating	% obtained	% of those receiving a given rating in 1938 who received the same or another rating in 1941				
		C	B—	B	B+	A
A	22.9%	.5%	1.1%	1.7%	3.5%	16.1%
B+	6.8	.7	1.0	.7	1.8	2.6
B	8.6	2.9	1.0	2.4	1.0	1.3
B—	8.0	3.9	1.1	1.0	1.5	.5
C	53.7	44.5	5.0	2.2	1.0	1.0
Ratings, 1941	100.0%	52.5%	9.2%	8.0%	8.8%	21.5%

sonable in view of the *lack* of any better instrument for the measurement of interests.

Somewhat apart, but nevertheless closely related to individual stability as revealed in test-retest correlations, are group trends in response to items. A rough measure of such changes was sought in the differences between means obtained on the serial tests. The differences were found to be statistically significant on only a few scales, with most of such differences tending toward a lower mean score at later testing.

The group became somewhat more heterogeneous with age, although many of the differences were not significant. This would, of course, contribute to the higher test-retest coefficients on later tests, as well as indicate a sharper delineation of interests at older age levels.

The subject of sex differences has arisen frequently in connection with vocational interests and should be considered here briefly. The presence of sex differences in vocational interests is rather generally accepted and is evident in the vocational choices of boys and girls. Authors of instruments for measurement of interests have tacitly implied the presence of significant differences by the construction of different tests for men and women. This fact, in itself, tends to inhibit the study of sex differences, since results from two separate tests, standardized on divergent samples, cannot be considered comparable, regardless of similarities in test form and procedure.

The administration of the Blank for Men to both boys and girls at the first testing has made possible some direct comparisons (Table VIII).

The girls obtained significantly higher mean scores on four scales, while the boys were higher on one scale. These differences tend to conform to established psychological sex differences, the girls being superior on verbal and social scales, and the boys on science scales. The boys, while significantly higher on one scale only, tend to exhibit a greater variability in response.

With the limitations in mind, it may be said that girls achieve somewhat higher mean scores, a greater percentage of "A" and "B+" ratings, and are slightly less variable than boys. Permanence coefficients attained by boys increase with age and show a fairly systematic development, with the length of the interval between tests as the variable. Such a development is not observed in the girls' scores, which show a tendency to be-

TABLE VIII

STANDARD ERROR OF THE DIFFERENCE BETWEEN THE MEAN SCORES OF BOYS AND GIRLS

Scale	r with M.–F.	Mean Boys (N = 64)	Mean Girls (N = 62)	Diff.[1]	Crit. Ratio: Diff. / σ Diff.
Physicist	.57	6218	6120	98	2.83
Lawyer	—.37	6149	6229	—80	3.94
Real Estate Salesman	—.32	6215	6288	—73.	3.56
Teacher	—.28	6222	6204	18	.81
Y. M. C. A. Secretary	—.40	6119	6098	21	.51
Office Worker	—.16	6331	6309	22	.96
C. P. A.	—.29	6214	6259	—45	4.20
Chemist	.64	6281	6155	126	6.63
Life Insurance Salesman	—.66	6175	6258	—83	3.99
Doctor	.33	6237	6237	0	0.00

[1] Differences marked with a negative sign are those in which the girls received a higher mean score than did the boys.

come less stable after high school graduation. Stability of letter ratings for both boys and girls is marked and is almost equal.

SUMMARY AND CONCLUSIONS

The purpose of this study was the investigation of the reliability and permanence of vocational interests of high school pupils, as measured over a four-year period of seriatim testing with the Strong Vocational Interest Blanks.

Reliability coefficients obtained by the correlation of odd and even halves of the tests ranged from .95 to .75, with an average coefficient of .87 for boys; they ranged from .96 to .78, with an average coefficient of .88 for girls. Fifteen of the vocational scales were found to have reliabilities of .85 or greater and consequently may be considered as suitable for use in individual predictions of interests.

Constancy of measured interests was investigated by correlations of scores on test and retest. The seriatim testing over a period of four years made possible the rotation of age and length of interval between tests. The correlations obtained with an increasingly greater interval between tests were progressively lower for both boys and girls, in spite of the increase in chronological age from test to test.

When scores obtained at varying age levels were correlated with length of interval held constant, the data for boys revealed small but reliable increases in permanence coefficients with increases in age. No such increase was found in a similar study of the girls' scores. A decline in permanence of measured interests after 18 years of age was indicated.

For purposes of comparison with published studies of interests at the college and older adult levels, an analysis of changes in letter ratings over a three-year period was made. It was found that 69 per cent of the ratings received by the boys at the first testing were identical with those received after the four-year interval. A similar finding was obtained for the data on girls; 66 per cent of the ratings obtained were duplicated after an interval of three years.

These results are similar to those obtained by investigators employing adult and college populations as samples. The conclusion is drawn that the inventories are approximately as adequate for use in the upper grades of the secondary school as they are for older persons, with respect to the reliability and permanence of measured interest scores. They indicate that, in practice, the measurement of interests by modern inventory methods may be extended to younger age levels with appropriate policies.

BIBLIOGRAPHY

1. Canning, Leslie B. "A Study of the Vocational Interests of Adolescent Boys by Repeated Testing with the Strong Vocational Interest Blank for Men." Unpublished M. A. thesis, University of California Library, Berkeley, Calif.
2. Canning, Leslie, Taylor, K. v. F., and Carter, H. D. "Permanence of Vocational Interests of High School Boys," *Journal of Educational Psychology*, XXXII (1941), 481–94.

3. Carter, H. D., and Jones, M. C. "Vocational Attitude Patterns in High-School Students," *Journal of Educational Psychology*, XXIX (1938), 321–34.
4. Carter, H. D., and Strong, E. K., Jr. "Sex Differences in Occupational Interests of High School Students," *Personnel Journal*, XII (1933), 166–75.
5. Carter, H. D., Taylor, K. v. F., and Canning, L. B. "Vocational Choices and Interest Test Scores of High School Students," *Journal of Psychology*, XI (1941), 297–306.
6. Fryer, Douglas. *The Measurement of Interests*. New York: Holt, 1931. Pp. xxxvi + 488.
7. Strong, E. K., Jr. "Permanence of Vocational Interests," *Journal of Educational Psychology*, XXV (1934), 336–44.
8. Strong, E. K., Jr. *Manual for Vocational Interest Blank for Women*. Stanford University Press, 1935. Pp. 15.
9. Strong, E. K., Jr. *Manual for Vocational Interest Blank for Men*. Stanford University Press, 1937. Pp. 16.
10. Strong, E. K., Jr. *Manual for Vocational Interest Blank for Women*. Stanford University Press, 1938. Pp. 16.
11. Taylor, Katherine van F. "A Study of the Reliability and Intensity of the Vocational Interests of High School Girls." Unpublished M.A. thesis, University of California Library, Berkeley, California.
12. Taylor, K. v.F., and Carter, H. D. "The Practical Applications of Objective Vocational Interest Data for the Vocational Counselor." To be published shortly in the *Journal of Consulting Psychology*.
13. Van Dusen, A. C. "Permanence of Vocational Interests," *Journal of Educational Psychology*, XXXI (1940), 401–24.

THE RELATIONSHIP BETWEEN PERFORMANCES ON EXAMINATIONS AND EMOTIONAL RESPONSES

WM. H. WAITE
State Teachers College
Milwaukee, Wisconsin

INTRODUCTION

Much attention has recently been given to the analysis of emotions in the great variety of situations in which emotional responses may be expected to occur. It is known that emotions, in general, cause hyperactive or hypoactive physiological reactions at the periphery of the body. These physiological changes afford additional convenient indices of certain aspects of human reaction and temperament, for it is thought that they may indicate how much a person is moved by his thoughts and feelings.

Numerous studies have been made of these physiological changes in breathing, pulse rate, galvanic skin reflex, and blood pressure under many types of stimulating conditions. These studies at first included the simultaneous observation of only one or two of these functions. The impossibility of making accurate comparisons of physiological changes where one kind of stimulus was used in studying one function and an entirely different kind of stimulus was used with the other resulted in further developments. Where no consideration was given to the type of stimulus or to the method of recording or treating the data, there were few significant and consistent findings. More comprehensive studies followed in which a large number of these physiological responses were recorded simultaneously during the application of standard stimuli. Laboratory apparatus also became standardized.

However, the criticism can be raised that the production of emotional reaction in an experimental laboratory is artificial and that the emotion, thus induced, cannot be designated as a genuine reaction. Moreover, the subject knows that emotion-producing stimuli are to be applied, and this anticipatory mental attitude tends to keep the attention of the subject upon himself while submitting to heroic and unusual measures.

The present study aims to overcome these objections by furnishing a stimulating situation that is more natural and genuine. During the examination situation the attention of the subject is diverted from himself, the apparatus, and the laboratory situation to the immediate results of his efforts upon the examination. A second important technique has been introduced into this study. Past experiments on the effects of examinations on physiological responses indicate that the measurements were made either before or after the examination or both. In the present experiment, by means of continuous records, a number of bodily changes were measured throughout the examination. This method gives to the investigator the important advantage of identifying the records of each physiological response with the corresponding part of the examination. Therefore, the physiological effects of mental attitudes during the experiment can be more readily classified and studied.

The measurement of physiological changes that accompany emotions or feelings has occupied the attention of a large number of individuals. As a result, numerous physiological measures have been developed. Among these measures are the pulse rate, the changes in blood pressure, the galvanic skin reflex, and changes in bodily resistance, the red and white corpuscle counts, the urine analysis for nitrogen, the fluctuations in the rate and the volume of respiration and the inspiration-expiration ratios, the postural changes, the hydrogen ion concentration of saliva, and the secretions of endocrine glands and peristalsis. This is a fairly representative but by no means a complete list. From this rather extensive list of measures have been selected those that have been used most extensively and which, for the purpose of this experiment, gave promise of the best results. A preliminary study and a careful review of the literature in the field led to a selection of three physiological measures. First, a continuous breathing record was obtained. From this record was secured the rate of respiration.

Second, a continuous record of the blood pressure furnished the pulse rate and blood pressure changes. Similarly, a continuous record of the galvanic skin reflex showed changes in bodily resistance and galvanic skin deflections.

The Apparatus and Interpretation

The physiological responses observed and recorded in the present experiment were: the level of resistance (palmar), the galvanic skin reflex, the pulse rate and blood pressure, and the respiration. The apparatus for recording these physiological changes consisted of a chair with a tablet arm inclined backward at an angle of approximately thirty degrees and equipped with a special leg rest for the left leg; and a Behavior Research Photopolygraph. The photopolygraph consisted of three types of recording apparatus for the determination of blood pressure and pulse rate changes, respiration changes, changes in the level of resistance, and galvanic skin reflex.

During the experiment the resistance and the galvanic skin reflex were recorded by a Behavior Research resistance box consisting of a Wheatstone bridge. One of the electrodes completing the circuit is a zinc disk 5/8" in diameter. This is rendered relatively nonpolarizable by covering it with a paste of kaolin (china clay) and a physiological solution (.9/10 per cent concentration) of zinc sulphate. These are enclosed in a bakelite cup with only the moist kaolin touching the subject's skin. This active electrode is applied to the palm of the left hand in the case of right-handed subjects and to the palm of the right hand for left-handed subjects. The second and indifferent electrode is oval in shape and of sheet zinc. It is two and one-half inches in width and three inches in length. The forearm to which the electrode is attached is vigorously massaged with Cambridge electrode jelly to reduce skin polarization. The contact surface of the electrode is also covered with a thin layer of the same substance.

Records of respiration are obtained by means of a thoracic pneumograph of a diaphramatic type, connected by a rubber tube with a Marey tambour. Inspiration changes the air pressure in the pneumograph and this change is transmitted through the rubber tube to the Marey tambour, which transmits the respiratory movements to the aluminum wire where it is recorded on the sensitized paper by means of shadow writing. The measurement of the rate of breathing is made on the breathing line. The rate of breathing is secured by a count of the number of peaks occurring in the breathing line during minute intervals. These are judged solely by inspection.

The pulse rate and blood pressure changes are obtained by using a continuous polygraphic method of recording. The heart rate is counted from the number of peaks caused by heart pulsations appearing on the blood pressure line. Measurement is made during every minute interval, a distance of eight and one-half millimeters upon the sensitized paper.

The frequency of these blood pressure fluctuations is calculated by the number of "waves" appearing on the blood pressure line during minute intervals. To be considered significant, a fluctuation in blood pressure had to be well rounded in shape and had to have an amplitude of at least one millimeter measured from the immediately preceding base line. The measurement of the amplitude of the fluctuation is secured by calculating in millimeters the distance from the fixed base line immediately preceding the rise to the maximum height of each peak. The duration of the fluctuation is measured from a preceding horizontal section to a following horizontal section or a return to the normal base line.

Procedure

The data of the present study were obtained from sixty boys of the University of Chicago High School. All of the subjects volunteered to participate in the experiment after the nature of the experiment had been explained to them.

On each of the subjects used, separate readings for each physiological function were obtained on three different occasions. The subjects were scheduled to report once a week on the same day and at the same time for three successive weeks. One-half hour was assigned to each subject. Twelve subjects were examined every day, sixty subjects a week.

Upon the subject's first entry into the laboratory a carefully planned attempt was made to put him at his ease and hasten adaptation. The method and ultimate purpose of

the experiment were explained to him. He was allowed to inspect the apparatus and to ask any questions that suggested themselves. After a desirable social relationship had been established, the subject was requested to sit in the reclining chair and to place his left leg along the special support. The trouser leg was rolled up and the pressure cuff applied just above the left ankle. A special foot rest help to eliminate any turning of the leg. The electrodes were then attached to the palm and forearm. The hand was then placed in the lap in a restful position. The subject was instructed to avoid all movement of the hand and fingers, if possible. The pneumograph was fastened around the chest by a leather belt and buckle. Subjects were instructed to wear the same type of clothing during their three visits to the laboratory, to notice particularly the position of the pneumograph on the chest, and to remember how it felt in order to assist the investigator in placing it.

After the apparatus had been attached, the electric motor which drove the film was started with the drive shaft out of gear. Moreover, the subject was placed with his back to the apparatus. As a further precaution a screen shut off from the subject's view all of the recording apparatus. This procedure prevented the subject from knowing when the actual recording was taking place and yet permitted the investigator to detect any movement on the part of the subject that might affect the records.

The subjects were instructed to remain as quiet as possible for a period of about ten minutes. They were warned not to fall asleep. After the apparatus was adjusted so that all parts were working satisfactorily, a five minute record was made. The purpose of this record was to secure the subjects' emotional reaction on their first visit to the laboratory.

The following week the subjects appeared for the second time at the laboratory. After the apparatus was adjusted and the motor started, the subjects were required to engage in simple cross-out exercises. The purpose of these exercises was to secure a condition somewhat comparable with the physical activities of a written examination situation. The subjects were advised that the exercises were of no particular consequence except for the purposes of concentration. A fifteen minute record was made during the second trial.

On the subjects' third visit to the laboratory, a written examination was substituted for the cross-out exercise. The examination consisted of three parts: number series, word completion, and artificial language, consisting of twenty questions, forty sentences, and twenty sentences respectively. The following instructions were given before the examination: "This is an examination of speed and accuracy. The results of this examination will determine your rank in a group of sixty boys. One copy of the results will be handed to your principal; another copy will be posted on the high school bulletin board. The examination consists of three parts, and you will have five minutes for each part. The first part is one page in length, the second and third parts, two pages in length. Do not begin until you receive the signal, 'ready, go.' You will be warned when only one minute remains in each part. When time is called, go to the next part. You are not expected to finish but answer as many questions as you can." This procedure was later adhered to during the conduct of the examination.

Since the same examination was given to subjects at various times of the day and on different days of the week, the subjects were pledged to refrain from discussing the nature of the questions with any other subject either before or after the examination. Each subject was interviewed immediately before the examination and, as far as could be ascertained, there was no collaboration among the subjects.

Although the examination was not based on accomplishment in academic subjects, every attempt was made to produce a natural and genuine examination situation. Some of the subjects were attending high school on scholarships and had scholastic prestige to maintain. Among the subjects, in general, scholarship was held at a high premium and competition was therefore very keen.

Variations in Physiological Responses During the Initial, Control, and Experimental Periods

The data for this part of the study were secured from the record sheets of sixty subjects.

The means of the raw scores during successive minutes of observation are calculated to show the general trends and levels of re-

sponse for each physiological measurement. The average score for the experimental period was higher than the corresponding average during the control period. This is true for each of the types of measures secured. In the case of log conductance, frequency of galvanic skin reflex, heart rate, respiration rate, and amplitude of fluctuations in blood pressure, the difference between the control and experimental periods in Table I is seen to be statistically significant.

A second general observation made from the graphs is that the scores during the experimental period show greater variability than those during the control period. It may be noted in addition that, except for the measure of the heart rate, the average scores during the initial period were above those for the corresponding minutes of the control period.

An examination of the individual types of measures shows some interesting results. For example, in log conductance the initial and control period records reveal that there is, under usual conditions, a gradual decrease in log conductance during the early minutes of a period of observations, followed by a gradual and regular increase that is maintained at least through the fifteenth minute. Beyond this period of time our observations did not extend. We find, in contrast to this picture during the experimental period, a regular pattern of increase and decrease in scores directly related to the five minute periods into which the examination was divided. After a high initial score, the experimental period displays a dropping off until the third minute. This decrease is followed by a sharp rise to the end of the sixth minute. After the sixth minute there is again a falling off followed by another rise to the tenth minute. This pattern repeats itself still again following the tenth minute. Thus, during this experimental trial there appears a definite pattern characterized first by a decrease, which is followed by a marked increase in conductance toward the end of each five minute period.

There was a general increased level of log conductance for each successive five minute period, the average log conductance for successive five minute periods being 133.92, 139.97, and 145.22. As was mentioned in the general observations, the level of log conductance during the experimental period was much greater than during the control period.

In general, the frequency of the galvanic skin reflex shows results similar to those for log conductance. There is, however, during the initial and control trials, no obvious tendency to a progressively lower score for the initial minutes of observation. The scores seem to increase fairly regularly and gradually throughout the period of observation. The initial minute of the experimental period is unusually high in score. This high score is

TABLE I

AVERAGE SCORE FOR THE VARIOUS CONDITIONS WITH THEIR STANDARD ERRORS AND THE CRITICAL RATIO FOR THE DIFFERENCE BETWEEN THE EXPERIMENTAL AND CONTROL PERIODS

Measurement	Initial Period	Control Period	Experimental Period	Difference between Control and Experimental periods.	Standard Error of difference	Critical Ratio of difference between Control and Experimental periods.
Log Conductance	134.76 ± 1.63	124.21 ± .91	139.70 ± .84	15.49	1.24	12.49
Frequency of Galvanic Skin Reflex	5.04 ± .29	3.93 ± .15	6.56 ± .14	2.63	.21	12.52
Respiration Rate	19.55 ± .18	19.22 ± .11	20.82 ± .10	1.60	.15	10.67
Heart Rate	80.60 ± .72	80.99 ± .37	85.15 ± .44	4.16	.58	7.17
Frequency of Fluctuations in Blood Pressure	.84 ± .09	.20 ± .03	.30 ± .03	.10	.04	2.50
Amplitude of Fluctuations in Blood Pressure	.36 ± .01	.08 ± .01	.14 ± .01	.06	.01	6.00
Duration of Fluctuations in Blood Pressure	2.53 ± .23	.64 ± .08	.88 ± .09	.24	.12	2.00

followed by a sharp decline during the following minute. At the fourth minute an emphatic increase is initiated, which lasts to the sixth minute. A sharp decline again occurs, and this decline in turn is followed by a pronounced increase to the tenth minute. A third decrease then takes place, followed by a marked increase to the end of the period. There is again a pattern of response clearly correlated with successive five minute examination periods. The average level of response, of course, is much higher for the experimental period and the general pattern is more irregular. As with log conductance, the averages for successive five minute periods of the experimental trial show a progressive increase, being 5.78, 6.72, and 7.18 respectively.

The scores in the rate of respiration in the initial and control trials show a general tendency to drop progressively during successive minutes of observation. During the experimental period a very marked difference is evident. The average score of 20.82 is noticeably higher than the general levels of the initial and control groups, and although during the first eleven minutes of the period there is an irregular decrease in rate, this is followed by a marked increase. During the last four minutes the breathing rate returns almost to the initial level of respiration for that trial. The presence of the examination seems to have produced a significant acceleration in the rate of breathing.

The data for the heart rate of the initial and control periods show a general tendency for the rate to increase progressively during the periods of observation. The heart rate during the experimental period shows considerable irregularity. The general level of response, determined by the average of the means, is much higher than in the case of the initial and control periods. A decrease during the second minute is followed by an increase, which is maintained to the fifth minute. A second and extensive decrease takes place, which is followed by an increase to the end of the ninth minute. A third and very slight decrease to the eleventh minute is followed by a rise, which is maintained to the fourteenth minute. The beginning of each part of the examination appears to be accompanied by a general retardation of the heart rate, followed by a gradual recovery and acceleration.

The frequency, amplitude, and duration of fluctuations in blood pressure show a very high level of response during the initial period. In contrast, the levels of responses during the control period are much lower. These responses for the experimental period are slightly higher than for the control period and show marked irregularity throughout. There is an observable similarity of pattern for the three measures in the experimental period.

The Effects of Successful, Average, and Unsuccessful Effort During a Short Period of Time on Certain Physiological Responses

Research relative to the general problem of the effects of examination difficulty on physiological responses has not been extensive and has in the main limited itself almost entirely to the measurement of physiological variations before and after the examination. However, experienced difficulty is a variable quantity that may express itself in varying degrees of intensity during the course of any one examination as the pupil interprets his chances of success or failure from the answers he makes. Moreover, it is possible to differentiate individuals who have taken an examination, on the basis of their performance as evidenced by their score on it. It then seems likely that those characterized by successful performance would have some awareness of their success, and conversely that those with poor performance would likewise have some knowledge of their difficulty. In order to get at the possible effects of such knowledge upon the physiological functions studied in this investigation, additional analyses of the data were made as described below.

As shown earlier, the examination was divided into three parts of five minute duration, and the physiological responses of the subject were secured during the writing of each part. The twelve best scores, the twelve average scores, and the twelve worst scores were then selected from each of the three parts and the corresponding concomitant physiological responses compared. By this approach an attempt was made to determine the effects of slight difficulty, average difficulty, and extreme difficulty upon the general trend of the physiological responses during a five minute period.

The graphs of log conductance show relative similarity in form both from period to period and from one type of performance to another. In each case the scores in general show a marked decrease during the initial part of the period, followed by an equally conspicuous rise to the end of the period. In the first period those who did best had, in general, a lower level of response in log conductance, while in the third period those making average scores show a lower level of response than those making the best or worst scores. A possible explanation of the lower level of response in the first period, for those making the best scores, might be a relatively high degree of self-confidence at the beginning of the examination period, while the difference found in the third period may perhaps be accounted for by greater indifference to the situation on the part of those making average scores. Those pupils making the best scores possibly were by then highly motivated by success, while those who made the worst scores were emotionally upset by the difficulty.

In the case of the scores of the frequency of galvanic skin reflex, the average group in the third period again shows a relatively low level of response. As already stated for log conductance, this low level of response may be due to the apparent indifference of the average group to the general situation.

In contrast to the results for log conductance and the galvanic skin reflex, there appear to be significant differences between those who made the best, average, and worst scores as indicated by the progressive changes in the rate of respiration. There appears to be a tendency for the rate of respiration to increase from period to period in the case of those making the best scores, to remain relatively constant in the case of those making the average scores, and for those making the worst scores to become progressively lower. This might be interpreted to mean that subjects who experienced success are more highly motivated and breathe more rapidly. Those who make average scores may be relatively stable and indifferent to the examination situation, and those students who make the worst score may experience a "blocking" that is reflected in the respiration rate.

In comparison with the other groups, those making the best scores appear to have a higher initial heart rate and to maintain a higher rate level throughout the respective examination periods. This is most marked, however, in the middle period. For the corresponding period of those making average scores, there is a much lower level of response, while the period for those making the worst scores has the lowest level of all.

The only significant difference in the graphs on the frequency, amplitude, and duration of fluctuations in blood pressure seems to be a relatively high level of response for those making the worst scores during the third period. We also find some tendency for the average scores to be lower throughout the three periods. In the duration of fluctuations, the best scores show a slightly higher level in the first period than either of the others. Irregularities and inconsistencies in the graphs for the blood pressure measurements render it difficult to make further observations.

The Effects of Continued Successful, Average, and Unsuccessful Effort on Certain Physiological Responses

The purpose of this part of the study is to examine and compare certain physiological responses accompanying continued success, continued average performance, and continued failure for the examination as a whole. As stated before, the examination was divided into three parts, and the physiological responses were obtained during the writing of each part. The best scores are the scores of the twelve subjects who made the highest marks on the three parts of the examination. Similarly, the average scores are the scores of the twelve subjects who made average marks on the three parts of the examination, and the worst scores are those of the twelve subjects who consistently made the lowest marks on the three parts of the examination.

A general characteristic of the graphs of log conductance is their relative similarity, with the exception that the average scores show very little of the usual pattern of change during the middle five minute period. Otherwise, the scores show a marked decrease during the earlier part of the period followed by rises at the sixth, tenth, and fifteenth minute. The average scores show a lower level of response than either the best or the worst scores. There is little difference in the behavior patterns of the levels of response

between pupils making the best and worst scores.

In the frequency of the galvanic skin reflex there appear to be no distinguishing characteristics between pupils making the best, average, and worst scores. The graphs have approximately the same initial level of response and are extremely irregular. All of the graphs show the common characteristic of marked increases at the sixth, tenth, and approximately the fifteenth minute.

In the breathing rate, however, significant differences between the three groups appear. Those making the worst scores have a low initial response and a gradual and regular decline until the thirteenth minute, from which point a partial recovery is made. The best and average scores show a high initial response. A gradual decline also takes place with a partial recovery occurring at the beginning of the thirteenth minute.

There are marked differences in the heart rates of those making the best, average, and worst scores. The general level of response for those making the best scores is extremely high and remains fairly constant as such throughout the fifteen minutes. The highest points on the graph occur at the fifth, tenth, and fourteenth minute. The level of response for the average scores is observably lower than for the best scores. The responses are more irregular than in the case of the other two classifications. Near the end of the period, the average scores become relatively stable and finish with a fairly high level of response. In the case of those making the worst scores, there is a lower level of response than is found in the other two types of scores. The trend is relatively stable.

The graphs of the frequency, amplitude, and duration of fluctuations in blood pressure are so irregular that they preclude significant observations. The frequency of fluctuations in the best scores shows a lower level of response and a more stable tendency than either of the other two measurements. In all three sets of graphs, the average and worst scores show a slight tendency toward a progressive decline, which is not shown by the graphs of the best scores.

Examination Results and Physiological Responses

As shown in the foregoing results, physiological changes occur in the body during the taking of an examination. There is a direct correspondence between the pattern of physiological change and the conditions of the examination. It has also been shown that, in general, those successful in the examination manifest an acceleration of some of the measurements of the physiological responses. In the section to follow, the physiological responses recorded during the writing of the examination are correlated with the examination marks to determine to what extent the quantitative results of the examination and the types of response expressed by the subjects are related. Table II indicates the amount of relationship between the two types of measures.

An inspection of Table II indicates that seven factors show positive correlation and four factors show negative correlation. In no case are their values sufficiently large to justify their use for suggestive relationship or predictive purposes. The largest coefficient is +.23, the correlation between the mean heart rate and success on the examination.

A further approach is made to the problem by relating the difference of the control and experimental means to the examination grades. The differences are calculated by subtracting the means of each physiological response for the control and experimental situation. The results are given in Table III.

Table II

The Relation Between Physiological Scores and Examination Grades

Physiological Responses	Coefficients of Correlation of Physiological Scores with Examination Grades
Log conductance	—.06 ± .09
Standard deviation of log conductance	+.11 ± .09
Frequency of galvanic skin reflex	+.03 ± .09
Standard deviation of galvanic skin reflex	+.11 ± .09
Heart rate	+.23 ± .08
Standard deviation of heart rate	—.05 ± .09
Frequency of fluctuations in blood pressure	+.02 ± .09
Amplitude of fluctuations in blood pressure	—.16 ± .09
Duration of fluctuations in blood pressure	—.04 ± .09
Respiration rate	+.06 ± .09
Standard deviation of respiration rate	+.08 ± .09

The data indicate that the values are again too small to be statistically significant.

In conclusion, the study failed to find any correlation either between the physiological responses recorded during the examination and the examination marks or between the physiological changes and the examination marks. However, emotional disturbances, due to the examination, are accompanied by physiological changes. The evidence also shows, in a general way, that some of the measures of the physiological responses are associated with successful and unsuccessful effort. Therefore, the lack of correlation between physiological measures and examination marks may be due to one or more of the following causes. There may be physiological changes associated with examination success that cannot be measured because of the lack of refined scientific apparatus or faulty technique. The pupils may have become so conditioned to examinations that minute individual physiological differences have become lost in the group. The examination, since it did not affect the scholastic rating of the subjects, perhaps failed to produce those physiological changes usually associated with individual success or failure.

Summary

Physiological psychologists have long recognized that beneath certain mental states of the individual, beneath such emotional reactions as anger, fear, anticipation, anxiety, disgust, etc., there are traceable definite physiological changes peculiar to the organism. However, a survey of the literature in the field shows that no previous study has attempted to take continuous records of physiological responses during the writing of examinations.

Sixty boys of the University of Chicago Laboratory Schools voluntarily agreed to participate in the experiment. Each subject visited the psychological laboratory on the same day and at the same time for three successive weeks. Physiological changes were recorded by a Behavior Research Photopolygraph consisting of three types of recording apparatus for the determination of changes in the level of resistance and the galvanic skin reflex, respiration changes, and changes in the blood pressure and pulse rate.

Three separate and distinct situations were studied: first, the initial period, the first visit of the subjects to the laboratory; second, the control period, the second visit of the subject to the laboratory a week later; and third, the experimental or examination situation.

Four types of study were undertaken. First, the effects of the initial, control, and experimental periods on the entire group of subjects and the statistical significance of the amount of change in the physiological measures between the control and experimental situation. Second, the effects of successful, average, and unsuccessful effort during a short period of time on certain physiological responses. Third, the effects of continued successful, average, and unsuccessful effort on certain physiological responses. Fourth, the relations between examination marks and physiological responses.

1. Statistically significant increases occurred during the examination period in log conductance, in the frequency of the galvanic skin reflex, in the respiration rate, in the heart rate, and in the amplitude of fluctuations in blood pressure.

2. Repetitious patterns approximately correlated with the five minute examination periods were most marked in the frequency of the galvanic skin reflex and somewhat less obvious in the measures of log conductance.

3. During the examination period, the rate of breathing was accelerated. The effect of the factors that caused this change seemed

Table III

The Relation Between Physiological Changes and Examination Grades

Physiological Changes Control to Experimental Period	Coefficients to Correlation of Physiological Variables with Examination Grades
Log conductance	−.20 ± .08
Standard deviation of log conductance	+.06 ± .09
Frequency of galvanic skin reflex	+.17 ± .09
Standard deviation of galvanic skin reflex	+.13 ± .09
Heart rate	+.15 ± .09
Standard deviation of heart rate	+.10 ± .09
Frequency of fluctuations in blood pressure	+.04 ± .09
Amplitude of fluctuations in blood pressure	−.11 ± .09
Duration of fluctuations in blood pressure	−.05 ± .09
Respiration rate	+.07 ± .09
Standard deviation of respiration rate	−.04 ± .09

most marked at the beginning and tended to decrease throughout most of the examination period.

4. The heart rate was very high during the first five minutes of the examination. A falling off occurred during the second five minutes, followed by a gradual recovery to the original level of response during the latter part of the examination.

5. The frequency, amplitude, and duration of fluctuations in blood pressure showed little consistency in pattern, and it is doubtful whether any conclusions may be drawn therefrom.

6. An attempt was made to discover the relation between the quality of the examination performance and the physiological responses. During a short examination period no significant differences in log conductance and the galvanic skin reflex occurred between those who made the best, average, and worst scores.

7. There was a tendency during a short examination period for the rate of respiration to increase from period to period in the case of those making the best scores, to remain relatively constant in the case of those making average scores, and to become progressively lower for those making the worst scores.

8. The measurements of the heart rate indicated that during a short examination period, those making the best scores appeared to have a higher initial heart rate and to maintain a higher rate level throughout the respective examination periods. This was most marked in the middle period. For the corresponding period of the average scores there was a much lower level of response, while that of the worst scores had the lowest level of all.

9. For those who made the best, average, and worst scores in continuous effort throughout the examination period, the outstanding differentiates were found in the breathing rate and the heart rate. In the breathing rate those making the worst scores had a low initial response and a gradual and regular decline until the thirteenth minute, from which point a partial recovery was made. The best and average scores showed a high initial response, followed by a gradual decline with a partial recovery occurring at the beginning of the thirteenth minute.

10. The general level of response for the heart rate during continued effort, for those making the best scores, was extremely high and remained fairly constant as such throughout the fifteen minutes. The level of response for those making average scores was observably lower than for the best scores, while for those making the worst scores there was a lower level of response than was found in the other two types of scores.

11. No significant correlations were found either between the physiological scores secured during the examination and the examination marks, or between the physiological changes for the control to the experimental period and the examination marks. Reasons have been advanced for this lack of relationship.

The findings of this experiment indicate that there is a general pattern of response for some physiological measures during examinations. There is also some suggestion that physiological responses of heart rate and respiration rate may be associated with success or failure on examinations of varying length, but the results are not conclusive.